

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

CARACTÉRISATION DU SITE COMMUN D'INTÉGRATION *KIS2* :  
IMPLICATION DE *MIR-106-363* DANS LES LEUCÉMIES DE TYPE T

THÈSE  
PRÉSENTÉE  
COMME EXIGENCE PARTIELLE  
DU DOCTORAT EN BIOLOGIE

PAR  
SÉVERINE LANDAIS

FÉVRIER 2007

UNIVERSITÉ DU QUÉBEC À MONTRÉAL  
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

*« We shall not cease from  
exploration, and the end of all our  
exploring will be to arrive where we  
started, and know the place for the  
first time. »*

**T.S Eliot**

*« Celui qui copie la nature est  
impuissant, celui qui l'interprète est  
ridicule, celui qui l'ignore n'est rien  
du tout. »*

**René Barjavel**

## REMERCIEMENTS

Mes premiers remerciements vont à mon directeur de thèse, le *Dr. Éric Rassart*, pour m'avoir donné l'opportunité de faire ma thèse dans son laboratoire. J'ai beaucoup appris, notamment grâce à la grande liberté qu'il m'a accordée quant à l'accomplissement de mon projet.

Je voudrais vivement remercier les membres de mon comité pour la correction de ma thèse et le jugement de mon travail, et particulièrement le *Dr. François Dragon* qui fut juge de chaque étape clé de ma thèse.

Je souhaite sincèrement remercier le *Dr. Robert Mamoun* qui m'a initiée à la recherche. Sa rigueur et son esprit scientifique seront toujours ma référence.

Je remercie mes collègues de travail côtoyés dans le laboratoire d'Éric, pour l'aide qu'ils m'ont apportée, de près ou de loin. J'y ai fait de belles rencontres ... et en particulier :

*Dr. Catherine Denicourt*, que j'ai eu la chance de connaître dans ma première année de thèse. Je la remercie pour tous ses conseils, mais aussi pour la passion qu'elle voue à la recherche et sa persévérance sans bornes qui furent pour moi très inspirantes.

*Véronique Voisin*, pour son soutien dans les moments difficiles, et pour tout ce que nous avons partagé lors de notre épopée québécoise, qui débuta un 17 Mai 2001 sur un quai en gare de Bordeaux.... Merci pour cette complicité qui fut par moments la nôtre. Nous y voici, finalement.... Souvenons-nous que rien ne dure, pas même la plus interminable des thèses !

*Fée-Ann McNabb*, pour sa simplicité, son sourire... Surtout je la remercie de nous avoir fait partager le bonheur de sa maternité... les belles joues d'Émile avaient un étrange pouvoir envoûtant, qui faisaient oublier tout stress !

*Paulina Salazar*, pour sa fraîcheur et son dynamisme, ainsi que son courage. Et aussi pour ses irrésistibles erreurs de français...

*Philippe Legault*, pour son assistance technique, et surtout ses blagues... Non, je voulais plutôt dire pour ses blagues, et surtout pour son assistance technique.... !

*Louis-Charles Levros*, pour sa gentillesse, et sa discrète présence qui trompait ma solitude durant les longues soirées et fins de semaine passées au laboratoire.

Je remercie tout particulièrement *Sébastien Landry*, avec qui j'ai eu le plaisir de collaborer. Nos grandes discussions scientifiques m'ont beaucoup aidée. Débattre sur mon travail m'aura manqué toute ma thèse, je regrette de ne pas t'avoir connu plus tôt ! Nous serons peut-être amenés à collaborer de nouveau, et cette fois-ci c'est sûr, on l'aura notre publication dans *Nature* !

Tous mes remerciements également aux gens des autres labos, qui m'ont apporté leur aide. Chaque petit geste et conseil, apparemment insignifiant, me nourriront pendant bien des années encore...

Je voudrais remercier mes amis, restés en France, pour avoir toujours été là malgré la distance. *Carine*, pour notre amitié de longue date, que le cœur emporte avec lui où que ce soit. *Guillaume*, pour nos belles correspondances, en riches diversions de mes tourmentes... Merci à Toi et *Céline* d'avoir été nos compagnons lors d'un fameux «road trip» en Gaspésie...

Je voudrais sincèrement (et fièrement), remercier mes parents pour leur confiance, leur soutien et leur ouverture d'esprit. Je remercie aussi ma famille pour ses encouragements. Malgré la distance, je n'oublie pas d'où je viens, et avec le temps, je le sais d'autant mieux.

Merci également à Claire et Sybil, pour tous ces agréables moments, qui ont su rendre ma fin de thèse plus humaine.

Mes ultimes remerciements sont pour *Fabrice*, avec tout l'amour possible. Il a su, souvent sans le savoir, tromper mes peurs et mes angoisses. Il m'a permis de rester en contact avec la vie, celle qui existe en dehors d'un laboratoire de recherches. Et au terme de ma thèse, je sais qu'il n'y a rien de plus précieux.

## TABLE DES MATIÈRES

REMERCIEMENTS.....	III
LISTE DES FIGURES.....	XI
LISTE DES TABLEAUX.....	XIV
LISTE DES ABRÉVIATIONS.....	XV
RÉSUMÉ.....	XVII
AVANT-PROPOS .....	XIX
METHODOLOGIE GENERALE.....	XXI

### CHAPITRE I

#### ÉTAT DES CONNAISSANCES.....1

1.1 L'UTILISATION DES RÉTROVIRUS EN TANT QU'OUTIL MOLÉCULAIRE DANS LA RECHERCHE D'ONCOGÈNES. ....	1
1.1.1 Généralité sur les rétrovirus.....	1
1.1.1.1 Composition d'un rétrovirus (Coffin 1997).....	1
1.1.1.1.1 La particule rétrovirale .....	1
1.1.1.1.2 Le génome rétroviral.....	3
1.1.1.1.3 Le cycle rétroviral (Coffin 1997).....	3
1.1.2 Effet de l'intégration de l'ADN rétroviral dans l'ADN de la cellule hôte .....	4
1.1.2.1 L'intégration : source de la pathogenèse rétrovirale.....	4
1.1.2.2 Les différents types de mutagenèse insertionnelle (Coffin 1997).....	6
1.1.2.2.1 L'insertion de promoteur .....	6
1.1.2.2.2 Activation transcriptionnelle .....	6
1.1.2.2.3 Élimination de séquences régulatrices du gène cellulaire.....	6
1.1.2.2.4 Interaction longue distance .....	7
1.1.2.2.5 Modification qualitative de la structure de la protéine .....	7

1.1.3 Les rétrovirus en tant qu'outils moléculaires pour la recherche d'oncogènes.....	9
1.1.3.1 Rétrovirus et cancérogenèse : historique.....	9
1.1.3.2 Rétrovirus non transformants et cancérogenèse : les sites communs d'intégration.....	10
1.1.3.2.1 La notion de site commun d'intégration comme hypothèse de travail .....	10
1.1.3.2.2 Recherche de SCI et découverte d'oncogènes par étiquetage rétroviral.....	11
1.1.3.2.3 Généralités sur le rétrovirus RadLV .....	12
1.2 ÉTAT DES CONNAISSANCES GÉNÉRAL SUR LES MicroARN .....	15
1.2.1 Biogenèse des microARN.....	15
1.2.1.1 Mécanisme de production des microARN (voir Fig. 1.5) .....	15
1.2.1.2 Organisation génomique des microARN.....	18
1.2.1.3 Profil d'expression des microARN .....	19
1.2.1.4 Régulation de l'expression des microARN.....	19
1.2.2 Fonctions des microARN .....	21
1.2.2.1 Inhibition de l'expression des gènes cibles et implication dans de nombreux processus biologiques. ....	21
1.2.2.1.1 Recherche des gènes cibles des microARN .....	22
1.2.2.1.2 Subtilités d'inhibition de l'expression des gènes par les miARN .....	23
1.2.2.2 L'exclusion mutuelle .....	26
1.2.2.3 Les microARN et la canalisation génétique.....	28
1.2.3 Les microARN et le cancer.....	29
1.2.3.1 Évidences de l'implication des microARN dans le cancer.....	29
1.2.3.2 MicroARN et leucémies .....	30
1.2.3.3 Typage et pronostic tumoral par profilage d'expression des microARN. ....	33

1.3	RÉGULATION POST-TRANSCRIPTIONNELLE DE L'EXPRESSION DES GÈNES PAR ÉPISSAGE ALTERNATIF .....	35
1.3.1	Mécanisme général de l'épissage .....	35
1.3.2	Épissage alternatif .....	40
1.3.2.1	Les séquences enhancer d'épissage (ESEs et ISEs).....	41
1.3.2.2	Les séquences silenciers d'épissage (ESSs et ISSs).....	41
1.3.2.3	Combinaison de facteurs dans la régulation de l'épissage.....	42

## CHAPITRE II

### DÉCOUVERTE ET CARACTÉRISATION DU SITE COMMUN

D'INTÉGRATION <i>KIS2</i> .....	46
---------------------------------	----

AVANT-PROPOS .....	46
--------------------	----

ARTICLE 1 : RADLV COMMON INTEGRATION AT THE <i>KIS2</i> LOCUS : SIMULTANEOUS OVEREXPRESSION OF A NOVEL NON-CODING RNA AND OF THE PROXIMAL <i>PHF6</i> GENE.....	47
---	----

2.1	ABSTRACT .....	48
2.2	INTRODUCTION .....	49
2.3	MATERIALS AND METHODS .....	51
2.3.1	Virus, mice and tumors.....	51
2.3.2	Genomic-DNA and RNA analysis .....	51
2.3.3	RadLV integration site amplification by I-PCR .....	52
2.3.4	5' and 3' RACE, CR-RT-PCR, and RT-PCR .....	53
2.3.5	cDNA cloning of <i>Phf6</i> .....	54
2.3.6	PHF6 localization in NIH3T3 cells.....	55
2.3.7	Sequence analysis.....	55
2.4	RESULTS .....	56
2.4.1	Integration pattern in RadLV-induced tumors.....	56



2.4.2	<i>Kis2</i> , a new common integration site.....	57
2.4.3	A novel gene is overexpressed in <i>Kis2</i> -rearranged tumors.....	59
2.4.4	Phf6 is also overexpressed in <i>Kis2</i> -rearranged tumors.....	61
2.5	DISCUSSION .....	64
2.6	REFERENCES.....	70
2.7	FIGURE LEGENDS.....	79

### CHAPITRE III

#### CARACTÉRISATION DES ARN NON-CODANT *KIS2* ET MISE EN ÉVIDENCE DE LEUR POTENTIEL ONCOGÈNE .....89

#### AVANT-PROPOS.....89

#### ARTICLE 2 : ONCOGENIC POTENTIAL OF THE *MIR-106-363* CLUSTER AND ITS IMPLICATION IN HUMAN T-CELL LEUKEMIA. .92

3.1	ABSTRACT.....	93
3.2	INTRODUCTION.....	94
3.3	MATERIALS AND METHODS .....	96
3.3.1	Cell culture.....	96
3.3.2	Human leukemias and normal individuals .....	96
3.3.3	Anchorage independence assay.....	96
3.3.4	miRNAs detection by Northern-Blot .....	97
3.3.5	miRNAs detection by padlock rolling circle amplification.....	97
3.3.6	RT-PCR on human RNA .....	98
3.3.7	5' and 3' RACE.....	99
3.3.8	Luciferase assay .....	99
3.3.9	Western-Blot.....	100
3.4	RESULTS .....	101
3.4.1	<i>Kis2</i> ncRNAs are the pri-miRNA for the <i>miR-106-363</i> cluster.....	101

3.4.2	The miRNAs from the <i>miR-106-363</i> cluster are overexpressed in <i>Kis2</i> -rearranged tumors.....	102
3.4.3	The <i>miR-106-363</i> cluster induces anchorage independence. ....	102
3.4.4	Expression analysis of <i>miR-106-363</i> cluster in human T-leukemias ...	103
3.4.5	Identification of potential target for the <i>miR-106-363</i> cluster .....	104
3.4.6	Endogenous MYLIP and RBP1-like are downregulated in NIH/3T3 by <i>miR-106-363</i> cluster.....	105
3.4.7	MYLIP and RBP1-like are downregulated in RadLV tumors overexpressing <i>miR-106-363</i> cluster.....	106
3.5	DISCUSSION .....	108
3.6	REFERENCES.....	112
3.7	FIGURE LEGENDS.....	115

## CHAPITRE IV

### IDENTIFICATION DES PARTENAIRES D'INTÉRACTION PROTÉIQUE DES ARN *KIS2* .....130

4.1	INTRODUCTION.....	130
4.2	MATÉRIEL ET MÉTHODES .....	133
4.2.1	Transcription <i>in vitro</i> (IVT).....	133
4.2.2	Fixation aux billes de Sepharose 4B activées au CNBr .....	133
4.2.3	Incubation avec les extraits protéiques.....	134
4.2.4	Séparation sur gel et séquençage des protéines .....	135
4.2.5	RNA Immunoprecipitation Assay (RIP) .....	135
4.2.6	Western-Blot.....	136
4.2.7	Transfection ADNc <i>Kis2</i> .....	136
4.2.8	Northern-blot.....	137
4.3	RÉSULTATS .....	138
4.3.1	Production des ARN <i>Kis2</i> par transcription <i>in vitro</i> et fixation aux billes	

de CnBr .....	138
4.3.2 Identification de protéines interagissant avec les ARN <i>Kis2</i> .....	138
4.3.3 Intérêt des protéines hnRNP A2/B1, hnRNP A3, PTB et KSRP .....	141
4.3.4 Le profil d'épissage des ARN <i>Kis2</i> observé dans les tumeurs diffère de celui observé dans les cellules NIH/3T3. ....	146
4.3.5 Analyse d'expression des protéines hnRNP A2/B1, hnRNP A3, PTB et KSRP dans les cellules BL/VL3 et NIH/3T3 .....	146
4.3.6 Les problèmes de confirmation des interactions des ARN <i>Kis2</i> avec les protéines .....	148
4.3.7 Implication pour le cluster <i>miR-106-363</i> .....	151
4.4 DISCUSSION .....	159
<b>DISCUSSION GÉNÉRALE ET PERSPECTIVES .....</b>	<b>164</b>
À propos des effets engendrés par les intégrations rétrovirales dans le génome hôte .....	164
À propos du gène <i>Kis2/miR-106-363</i> .....	166
À propos du sens de l'épissage alternatif du gène <i>Kis2</i> .....	170
<b>RÉFÉRENCES BIBLIOGRAPHIQUES .....</b>	<b>172</b>

## LISTE DES FIGURES

<b>Figure 1.1:</b> Structure générale des rétrovirus. ....	2
<b>Figure 1.2:</b> Le cycle rétroviral.....	5
<b>Figure 1.3:</b> Les différents types de mutagenèses insertionnelles .....	8
<b>Figure 1.4:</b> Technique de la PCR inversée (Li <i>et al.</i> 1999).....	14
<b>Figure 1.5:</b> Production des miARN (extrait de Kim <i>et al.</i> 2006). ....	17
<b>Figure 1.6:</b> Exemple de structures tige-boucle des pre-miARN (extrait de Bartel <i>et al.</i> 2004).. ....	20
<b>Figure 1.7:</b> Exemple d'hybridation d'un miARN avec sa séquence cible (extrait de Rajewsky <i>et al.</i> 2006).....	25
<b>Figure 1.8:</b> (A) Exemple du rôle de l'exclusion mutuelle dans le développement (extrait de Hornstein 2006). (B) Illustration des FFL cohérent et incohérent.....	27
<b>Figure 1.9:</b> Illustration des mécanismes moléculaires impliquant les miARN dans le cancer (extrait de Zhang <i>et al.</i> 2006). ....	31
<b>Figure 1.10:</b> Séquences consensus d'épissage (extrait de «Molecular biology of the cell», Alberts 2002).....	36
<b>Figure 1.11:</b> Mécanisme général de l'épissage (extrait de Black <i>et al.</i> 2003) .....	37
<b>Figure 1.12:</b> La définition lors de l'épissage (Extrait de Berget <i>et al.</i> 1995).....	39
<b>Figure 1.13:</b> Épissage alternatif.....	43
<b>Figure 1.14:</b> Épissage alternatif de l'exon N1 du gène c-src (extrait de Black <i>et al.</i> 2003).. ....	45
<b>Figure 2.1:</b> Positions of RadLV integrations in <i>c-myc</i> , <i>Pim1</i> and <i>Notch1</i> loci.. ....	82
<b>Figure 2.2:</b> Southern blot analysis of some RadLV-induced leukemias.....	83

<b>Figure 2.3:</b> (A) Northern blot analysis of the <i>Kis2</i> RNA species and schematic representation of their structure. (B) Semi-quantitative analysis by nested RT-PCR of the expression level of <i>Kis2</i> transcripts in normal mouse tissues. ....	84
<b>Figure 2.4:</b> Analysis of the <i>Kis2</i> sequence conservation among species.....	85
<b>Figure 2.5:</b> Analysis of <i>Phf6</i> gene expression in tumors and normal mouse tissues. (A) <i>Kis2</i> -rearranged (T1 to T6) and non-rearranged (NR) tumors by RT-PCR. (B) Nested RT-PCR on a panel of normal tissues .....	86
<b>Figure 2.6:</b> Schematic representation of all the <i>Phf6</i> transcripts.....	87
<b>Figure 2.7:</b> Sub-cellular localization of the PHF6 protein from transcript 2 and 7....	88
<b>Figure 3.0:</b> Situation des miARN du cluster miR-106-363 dans la région <i>Kis2</i> .....	91
<b>Figure 3.1:</b> (A) Genomic map of the <i>Kis2</i> locus on mouse chromosome X..... (B) miRNAs organization of the <i>miR-17-92</i> and <i>miR-106-363</i> clusters, and sequence comparison between miRNAs from each cluster..... (C) <i>Kis2</i> ncRNAs species identified as pri-miRNA for <i>miR-106-363</i> cluster.....	118 119 120
<b>Figure 3.2:</b> (A) Expression analysis of miRNAs from the <i>miR-106-363</i> cluster in RadLV-induced tumors..... (B) Quantification of signal relative to Northern-Blot in (A)..... (C) Transfection of each of miR-106a, miR-19b-2, miR-20b and miR- 92-2 induces anchorage independence in a soft agar assay..... (D) Colony counting from (C) .....	121 122 123 123
<b>Figure 3.3:</b> (A) Expression analysis of the <i>miR-106-363</i> cluster in human T- leukemias..... (B) Quantification of signal observed in (A).....	124 125
<b>Figure 3.4:</b> Schematization of Hipk3, Mylip, and Rbp1-like 3'UTR. ....	116
<b>Figure 3.5:</b> (A) Analysis of different 3'UTR in a luciferase assay..... (B) Western Blot of the RBP1-like and MYLIP protein levels in NIH/3T3 cells transfected with either miR-106a, miR-19, miR-92, miR-20b.....	127 128

(C) Western blot of MYLIP, RBP1-like and RB protein levels in tumors overexpressing or not the <i>miR-106-363</i> cluster .....	129
<b>Figure 4.1:</b> Transcription <i>in vitro</i> des ARN <i>Kis2</i> .....	139
<b>Figure 4.2:</b> Séparation des protéines sur gel de polyacrylamide (gradient).....	142
<b>Figure 4.3</b> Comparaison entre les ARN <i>Kis2</i> observés dans les tumeurs et les ARN <i>Kis2</i> dans les cellules NIH/3T3.....	144
<b>Figure 4.4</b> Combinaison d'inhibition/ activation d'épissage à l'origine des transcrits <i>Kis2</i> .....	145
<b>Figure 4.5</b> Western-blot de comparaison entre cellules NIH/3T3 et BL/VL3 pour les protéines hnRNP A2/B1, hnRNP A3, KSRP et PTB.....	147
<b>Figure 4.6</b> Résultats du RIP spécifique de PTB et hnRNP A2/B1.....	150
<b>Figure 4.7</b> Prédiction de la structure de l'ARN <i>Kis2</i> t1, du pri-miARN complet correspondant au transcrit <i>Kis2</i> t1.. .....	153
<b>Figure 4.8</b> Prédiction de la structure de l'ARN <i>Kis2</i> t2, du pri-miARN complet correspondant au transcrit <i>Kis2</i> t2. ....	154
<b>Figure 4.9</b> Prédiction de la structure de l'ARN <i>Kis2</i> t3, du pri-miARN complet correspondant au transcrit <i>Kis2</i> t3.. .....	155
<b>Figure 4.10</b> Prédiction de la structure de l'ARN <i>Kis2</i> t4, du pri-miARN complet correspondant au transcrit <i>Kis2</i> t4.. .....	156
<b>Figure 4.11</b> Prédiction de la structure de l'ARN <i>Kis2</i> t5, du pri-miARN complet correspondant au transcrit <i>Kis2</i> t5.. .....	157
<b>Figure 4.12</b> Prédiction de la structure de l'ARN <i>Kis2</i> t6, du pri-miARN complet correspondant au transcrit <i>Kis2</i> t6. ....	158

## LISTE DES TABLEAUX

<b>Tableau 1.1</b> Exemple de recherche d'oncogène obtenue par étiquetage rétroviral à haut débit (Mikkers <i>et al.</i> 2002) .....	13
<b>Tableau 1.2</b> Différents outils informatiques pour la prédiction de cibles de miARN .....	24
<b>Tableau 1.3</b> Exemple de miARN impliqués dans divers types de cancers (extrait de Garzon <i>et al.</i> 2006) .....	32
<b>Table 3.1</b> <i>miR-106-363</i> target genes predicted and tested in our luciferase assay. .	117
<b>Tableau 4.1</b> Résumé du séquençage par spectroscopie de masse des bandes sélectionnées sur gel .....	140

## LISTE DES ABRÉVIATIONS

ADN	Acide désoxyribonucléique
ADNc	ADN complémentaire
ALV	<i>Avian Leukosis Virus</i>
AREs	<i>AU Rich Elements</i>
ARN	Acide ribonucléique
ARNm	ARN messenger
ARNnc	ARN non codant
ATP	Adénosine Tri-Phosphate
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
CMV	Cyto Megalo Virus
Da	Dalton
DCS	<i>Downstream Control Sequence</i>
DTT	Dithiothreitol
ESE	<i>Exon Splicing Enhancer</i>
ESS	<i>Exon Splicing Silencer</i>
FFL	<i>Feed-Forward Loop</i>
FT	Facteur de Transcription
hnRNP	<i>heterogeneous RiboNuclear Protein</i>
ISE	<i>Intron Splicing Enhancers</i>
ISS	<i>Intron Splicing Silencer</i>
IVT	<i>In Vitro Transcription</i>
Kb	Kilo bases
kDa	kiloDalton
<i>Kis2</i>	<i>Kaplan Integration Site 2</i>
KSRP	<i>K-Homology Splicing Regulatory Protein</i>



LLC	Leucémies Lymphocytiques Chroniques
LTR	<i>Long Terminal Repeat</i>
miARN	microARN
MoMuLV	Moloney MuLV
MuLV	<i>Murine Leukemia Viruse</i>
pb	paires de bases
PBMCs	<i>Peripheral Blood Mononuclear Cells</i>
PBS	<i>Phosphate Buffer Saline</i>
PCR	<i>Polymerase Chain Reaction</i>
PMSF	PhenylMethaneSulphonylFluoride
Pre-ARNm	ARNm non épissé
PSF	<i>PTB associated Splicing Factor</i>
PTB	<i>Polypyrimidine Track Binding protein</i>
RadLV	<i>Radiation Leukemia Virus</i>
RBP	<i>RNA Binding Proteins</i>
RIP	<i>RNA ImmunoPrecipitation</i>
RISC	<i>RNA-induced silencing complex</i>
RNP	<i>Ribo-Nucleo Particules</i>
RT	<i>Reverse Transcriptase</i>
RT-PCR	<i>Reverse Transcriptase PCR</i>
SAE	Site Accepteur d'Épissage
SCI	Site Commun d'Intégration
SDE	Site Donneur d'Épissage
SDS	Sodium Dodecyl Sulfate
siRNA	<i>silencing RiboNucleic Acid</i>
snRNP	<i>small nuclear RiboNucleoProtein</i>
SR	<i>Arginin and Serine rich proteins</i>
UTR	<i>UnTranslated Region</i>
VIH	Virus de l'immunodéficience Humaine

## RÉSUMÉ

L'objectif de ma thèse était de trouver de nouveaux oncogènes impliqués dans les leucémies de type T, en utilisant la méthode d'étiquetage rétroviral. Pour induire des leucémies de type T chez la souris, nous utilisons le rétrovirus RadLV qui induit la maladie en un temps de latence très court.

La première étape de mon travail a consisté en la recherche de sites communs d'intégration (SCIs) de RadLV dans des thymomes de souris induits par ce rétrovirus. Nous avons trouvé un nouveau SCI que l'on a appelé *Kis2*, situé sur le chromosome X et impliqué dans 11% des tumeurs analysées. Après avoir répertorié les gènes du locus, nous avons montré que les intégrations de RadLV au niveau du locus *Kis2* provoquaient la surexpression de deux gènes. Le premier, situé très proche du SCI, correspond à un ARN non-codant que l'on a baptisé *Kis2*. Le deuxième, situé 300 kb en amont, correspond au gène *Phf6*, récemment impliqué dans un syndrome de retard mental. Nous avons donc mis en évidence que les intégrations rétrovirales pouvaient activer plusieurs gènes à la fois, et ce sur de longues distances.

La deuxième étape de mon travail a consisté en la caractérisation du gène *Kis2*. Les intégrations rétrovirales provoquent la surexpression de cinq transcrits de ce gène, mettant en évidence un profil d'épissage complexe. Son caractère non-codant a rendu ce gène énigmatique, jusqu'à ce que l'on découvre en 3' du gène la présence du groupe de microARN (miARN) *miR-106-363*. Nous avons pu montrer qu'en définitive les ARN *Kis2* correspondaient au pri-miARN de *miR-106-363*, qui se compose de six miARN (*miR-106a*, *miR-18b*, *miR-20b*, *miR-19b*, *miR-92-2* et *miR-363*). La surexpression de ces pri-miARN dans les tumeurs de souris conduit à l'accumulation de *miR-106a*, *miR-20b*, *miR-19b* et *miR-92-2*. Dans 46% des leucémies de type T humaines, le pri-miARN *miR-106-363* est également surexprimé, conduisant à une surproduction de *miR-19b* et *miR-92-2*, et parfois de *miR-106a*. Nous montrons donc l'implication de *miR-106-363* dans les leucémies humaines, et également l'existence d'une régulation post-transcriptionnelle de ce groupe dont tous les miARN ne sont pas produits en même temps et dans les mêmes quantités.

De façon intéressante, *miR-106-363* est homologue au groupe *miR-17-92*, connu pour être impliqué entre autres dans des lymphomes de type B. Pour apprécier le potentiel oncogénique de *miR-106-363*, nous avons réalisé un test d'indépendance d'ancrage sur des cellules NIH3T3. Les résultats indiquent que *miR-106a*, *miR-20b*, *miR-19b* et *miR-92-2* sont capables d'induire la formation de colonies.

Nous avons par la suite procédé à l'identification de gènes cibles de ces miARN. Parmi une sélection de prédictions informatiques communes à *miR-106a/20b*, *miR-19b* et *miR-92*, nous avons choisi les gènes à caractère suppresseur de tumeur. Nous montrons que les gènes *Myli1*, *Rbp1-like* et possiblement *Hipk3* sont ciblés par ces miARN, et que leur

expression est inversement corrélée à celle de *miR-106-363* dans les tumeurs de souris. Nos résultats constituent de nouveaux indices quant aux mécanismes oncogéniques relatifs à la surexpression de *miR-106-363*.

Le dernier volet de ma thèse concerne l'identification de protéines interagissant avec les ARN *Kis2*. Ce travail était destiné à nous renseigner sur le rôle de ces ARN, alors que nous n'avions pas connaissance de *miR-106-363*. J'ai identifié par spectrométrie de masse trois protéines candidates : hnRNP A2/B1, hnRNP A3, KSRP, et possiblement PSF. Le rôle de ces protéines orienté vers la régulation de l'épissage m'a conduit à formuler l'hypothèse de l'implication de ces protéines dans la production des différents transcrits du gène *Kis2*. Finalement, ce profil d'épissage complexe pourrait être relié à l'expression différentielle des miARN de *miR-106-363*, et constituer un élément de régulation post-transcriptionnelle de l'expression des miARN. Les interactions de ces protéines avec les ARN *Kis2* restent toutefois à confirmer *in vivo*.

Mots clé : leucémie, étiquetage rétroviral, oncogène, microARN.

## AVANT-PROPOS

En 1976, les travaux de Varmus et Bishop donnent naissance à l'ère de la cancérologie moderne. Pour la première fois, des chercheurs mettaient en évidence que la composante virale responsable du sarcome de Rous chez le poulet n'était autre qu'un gène cellulaire (Stéhelin 1976). Le mal prenait donc sa source directement dans nos cellules, et le cancer apparut dès lors comme une maladie de «gènes». Ces gènes au cœur du cancer sont les oncogènes et les suppresseurs tumeurs. Avec cette découverte majeure, qui fut gratifiée du prix Nobel de médecine en 1989, les chercheurs pensaient avoir gagné la bataille : il suffisait de débusquer les oncogènes et suppresseurs tumeurs pour faire du cancer une maladie presque anodine. Un quart de siècle et quelques gènes plus tard, force est d'admettre que l'adversaire est plus complexe qu'il n'y paraissait, et le cancer fait toujours des ravages. Selon le World Cancer Report, environ 10 millions de cancers ont été déclarés en l'an 2000 à l'échelle mondiale, et 6,2 millions de personnes sont mortes de leur maladie.

Le cancer représente donc toujours un grave problème de santé publique et un coût énorme pour la société. En dépit des grands progrès réalisés en matière de compréhension du cancer et de traitements développés, la prévention et le dépistage précoce restent encore les meilleures armes. En fait plus nos connaissances en biologie du cancer progressent, plus elles dévoilent une étourdissante complexité. En effet, les outils de biologie moléculaire de pointe tels que les micropuces ont permis l'analyse du transcriptome d'une cellule cancéreuse, et ont révélé des changements d'expression génique majeurs par rapport à une cellule normale. Paradoxalement, il est admis que seul un petit nombre de voies principales de transformation cellulaire existent, et bon nombre de gènes clés qui s'y rattachent ont déjà été découverts (Hahn 2002). Cependant, le ciblage thérapeutique de ces gènes ne suffit pas à résoudre tous les cancers. De plus, une cellule cancéreuse, aussi aberrante soit-elle, reste encore très semblable à une cellule normale, ce qui rend les traitements administrés très toxiques pour le

patient. Certains chercheurs pensent que le cancer est une maladie trop complexe pour ne considérer qu'un petit nombre de gènes à la fois, aussi importants soient-ils. La clé pour comprendre et déjouer la complexité du cancer réside peut-être dans une considération plus globale des changements qui surviennent dans une cellule cancéreuse. Pour cela, il apparaît essentiel non seulement de trouver le maximum de gènes impliqués dans le cancer, mais aussi de déjouer les connexions qui existent entre eux.

Dans notre laboratoire, nous sommes intéressés par les gènes impliqués dans la leucémie. À notre époque, cette maladie représente environ 550 000 cas par année, avec un taux de mortalité élevé, en particulier chez les enfants. Nous espérons que nos recherches, aussi modestes soient-elles, constitueront une étape supplémentaire et nécessaire à la compréhension de la leucémie, et également des autres cancers.

## **MÉTHODOLOGIE GÉNÉRALE**

### **INJECTION DU RÉTROVIRUS RADLV À DES SOURIS NOUVEAUX NÉES**



### **ANALYSE DES TUMEURS : RECHERCHE DES SITES D'INTÉGRATION RÉTROVIRALE**

Une fois que les souris développent des tumeurs, les thymus sont prélevés pour en extraire l'ADN. Les intégrations rétrovirales sont trouvées grâce à la méthode de PCR inverse. Les amplicons sont clonés puis séquencés. La position des intégrations est alors déterminée par comparaison de séquences dans les banques de données telles que [ensembl.org](http://ensembl.org).



### **LES SITES D'INTÉGRATION IDENTIFIÉS SONT-ILS DES SITES COMMUNS D'INTÉGRATION DU RÉTROVIRUS ?**

Chaque site d'intégration précédemment identifié est par la suite étudié afin de déterminer s'il s'agit d'un site commun d'intégration, c'est-à-dire un locus au niveau duquel on retrouve des intégrations dans des tumeurs d'origines différentes. Ceci est réalisé au moyen de Southern blot, en combinaison avec la PCR. Cette étape capitale constitue la base de notre travail.



### **RECHERCHE DE GÈNES ASSOCIÉS AUX SITES COMMUNS D'INTÉGRATIONS = ONCOGÈNES POTENTIELS**

Une fois qu'un site commun d'intégration est identifié, les gènes du locus dont l'expression est modifiée par la présence du rétrovirus sont recherchés. Ces gènes représentent de potentiels oncogènes, et seront par la suite caractérisés.

## **CHAPITRE I**

### **ÉTAT DES CONNAISSANCES**

#### **1.1 L'UTILISATION DES RÉTROVIRUS EN TANT QU'OUTIL MOLECULAIRE DANS LA RECHERCHE D'ONCOGENES.**

##### **1.1.1 GÉNÉRALITÉ SUR LES RÉTROVIRUS**

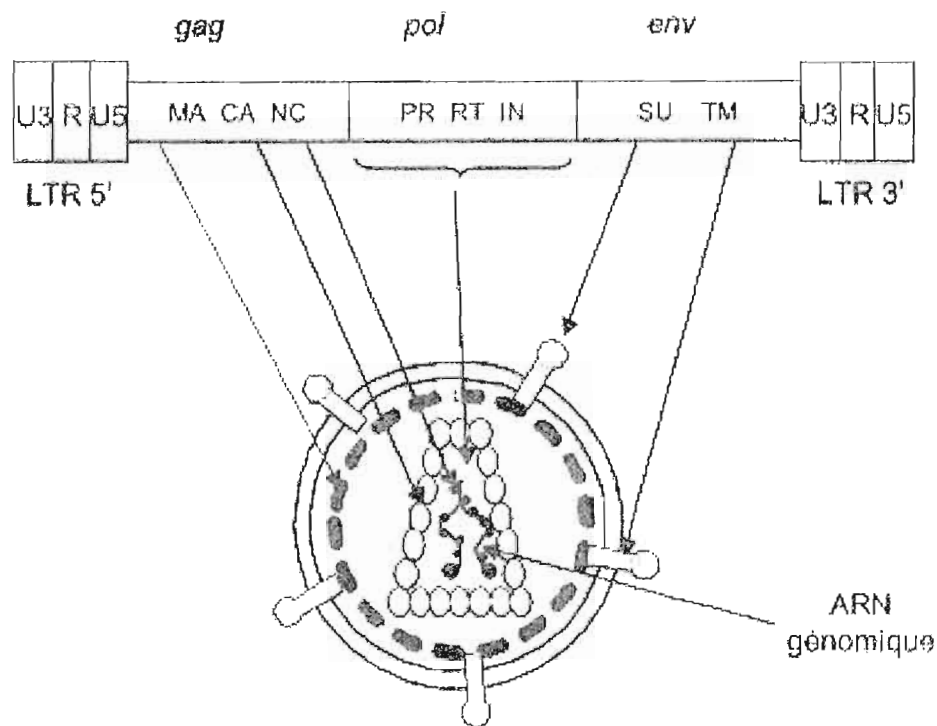
Le nom rétrovirus vient de la propriété de ces virus à ARN de rétrotranscrire leur ARN en ADN, qui s'intègre ensuite dans le génome cellulaire. Les rétrovirus existent dans de nombreuses espèces y compris chez l'homme. Ils provoquent différents types de maladies : des cancers ou des leucémies, des maladies auto-immunes, des maladies dégénératives à évolution lente, des maladies du système immunitaire se traduisant par une immunodépression profonde.

##### **1.1.1.1 Composition d'un rétrovirus**

###### **1.1.1.1.1 La particule rétrovirale (Fig. 1.1):**

D'une façon générale, la particule rétrovirale se compose successivement de

- l'enveloppe où est ancrée la glycoprotéine de surface (SU), associée à une protéine transmembranaire (TM),



**Figure 1.1 Structure générale des rétrovirus (Coffin *et al.* 1997).** Le génome rétroviral se compose de trois gènes principaux : *gag* codant pour les protéines de la matrice (MA), les protéines de la capside (CA), et les protéines de la nucléocapside (NC), *pol* codant pour la protéase (PR), la rétrotranscriptase (RT), la RNase H (RH) et l'intégrase (IN), et *env* codant pour les protéines de l'enveloppe (SU et TM). En 5' et 3' du génome rétroviral se trouvent les LTRs (Long terminal repeat) divisés en trois régions distinctes (U3, R, U5).



- une couche de protéine de matrice (MA),
- une couche de protéine de capsid (CA), qui renferme deux copies d'ARN génomique simple brin polyadénylés à polarité positive associées à des protéines de nucléocapside (NC),
- les enzymes virales : protéase (PR), rétrotranscriptase (RT), RnaseH (RH) et intégrase (IN)

#### 1.1.1.1.2 Le génome rétroviral

Il y a trois gènes rétroviraux principaux communs à tous les rétrovirus (Fig. 1.1): gag (codant pour MA, CA et NC), pro-pol codant pour PR, RT, RH et IN, et env codant pour SU et TM. Les rétrovirus possédant d'autres gènes qui codent pour des protéines dites accessoires sont qualifiés de rétrovirus complexes. De plus, certains rétrovirus possèdent dans leur génome un oncogène d'origine cellulaire, comme c'est le cas du virus du sarcome de Rous (Stéhelin *et al.* 1976). Ces rétrovirus sont qualifiés de transformants et induisent des tumeurs en un temps de latence très court. Les MuLVs (*Murine Leukemia Viruses*) que nous utilisons comme outils se situent dans la catégorie des rétrovirus simples et non transformants. Au niveau des extrémités 5' et 3' du rétrovirus se trouvent les LTRs (*Long Terminal Repeat*), qui se composent de séquences promotrices et enhancers puissantes responsable de la transcription des ARN rétroviraux.

#### 1.1.1.1.3 Le cycle rétroviral

Pour comprendre l'utilité des rétrovirus dans la recherche d'oncogènes, il faut décrire brièvement le cycle rétroviral (Fig. 1.2). Lors de l'infection d'une cellule, le rétrovirus s'accroche à un récepteur cellulaire spécifique par l'intermédiaire de sa protéine de surface. L'enveloppe virale et la membrane cellulaire fusionnent, et la nucléocapside pénètre dans la cellule. L'ARN rétroviral est rétrotranscrit en double brin d'ADN grâce à la transcriptase inverse. La RNase H intervient pour la dégradation de l'ARN lors du processus de rétrotranscription. L'ADN viral est ensuite transporté dans le noyau, où il sera intégré au

génomique de la cellule grâce à l'intégrase. Il prend alors le nom de provirus. Les gènes rétroviraux sont transcrits par l'ARN polymérase II cellulaire. Des copies complètes d'ARN du génome parental sont générées, et donnent naissance par épissage alternatif aux différents ARNm des protéines du virion. Ces ARNm sont traduits dans le cytoplasme cellulaire. Une partie de ces copies complètes d'ARN rétroviral sera encapsidée pour constituer le génome des virus produits. Les différents constituants viraux s'assemblent au niveau de la membrane cellulaire, et la particule virale est relarguée par bourgeonnement. La particule relarguée est immature, et subit par la suite un processus de maturation.

## **1.1.2 EFFET DE L'INTÉGRATION DE L'ADN RÉTROVIRAL DANS L'ADN DE LA CELLULE HÔTE.**

### **1.1.2.1 L'intégration : source de la pathogenèse rétrovirale**

C'est l'étape d'intégration du génome rétroviral dans l'ADN cellulaire qui est à l'origine de la pathogenèse des rétrovirus non transformants. Le fait que le génome rétroviral soit intégré dans le génome cellulaire peut avoir diverses conséquences sur la physiologie de la cellule. D'une part, la capacité des rétrovirus à s'intégrer quasiment au hasard dans l'ADN en fait des mutagènes potentiels puissants. Ceci est qualifié de mutagenèse insertionnelle. Cette mutagenèse peut rester silencieuse, toucher un locus affectant un caractère secondaire (couleur du pelage), mettre en jeu la viabilité de l'animal (mutation létale récessive), affecter des fonctions importantes mais non critiques pour la survie (anémie) ou enfin modifier des gènes potentiellement oncogéniques, ce qui peut conduire à la formation de tumeurs. D'autre part, la présence au sein du génome viral des séquences cis-régulatrices puissantes que sont les LTR explique que l'insertion des rétrovirus à proximité de gènes cellulaires puisse entraîner une dérégulation de leur expression.

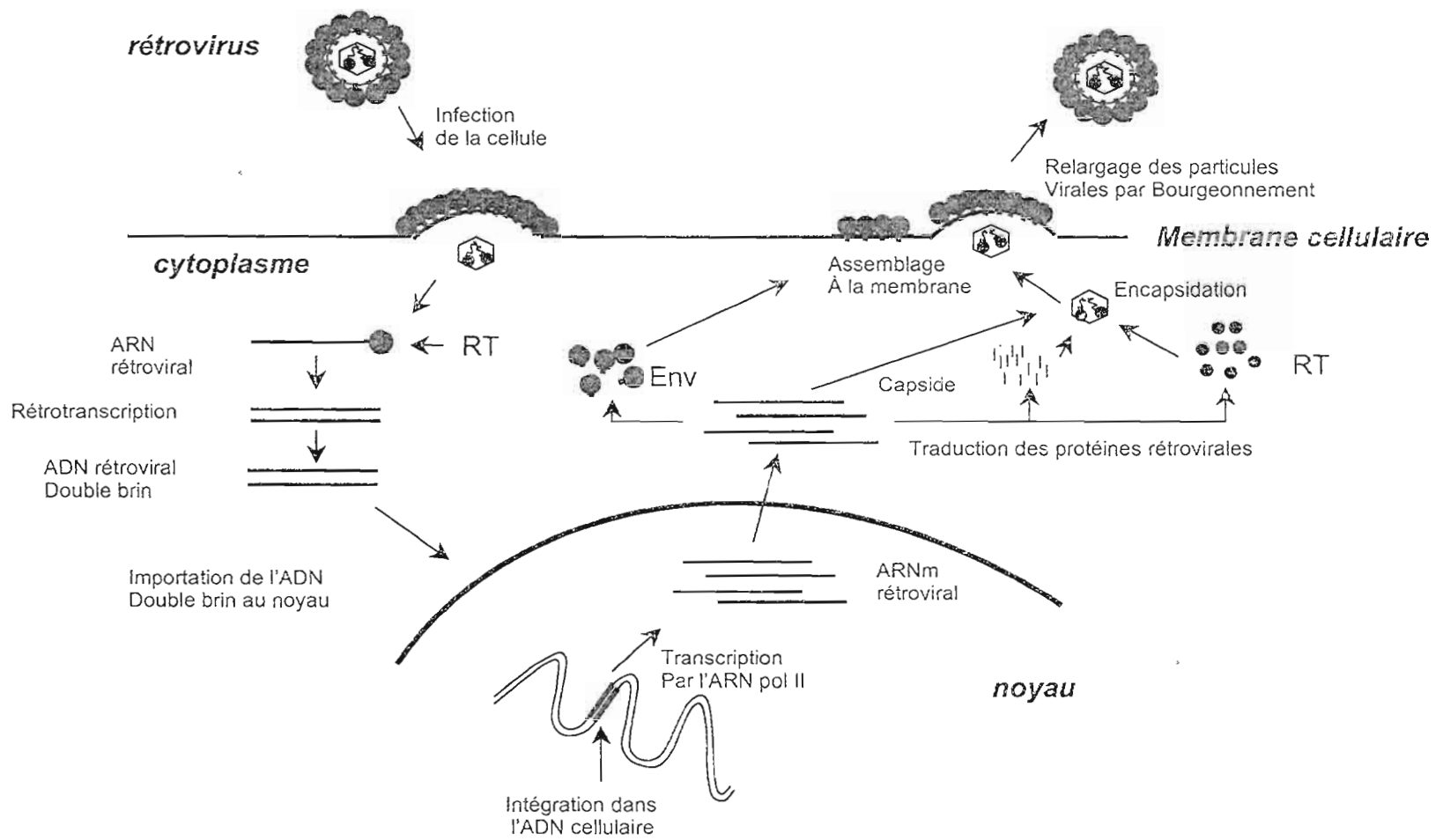


Figure 1.2 Le cycle rétroviral (voir texte page 3 pour explication)

Enfin, si le virus comporte un gène dont le produit intervient dans la physiologie cellulaire (cas des rétrovirus transformants), il peut être dangereux pour la cellule: l'intégration de ce gène supplémentaire et son expression souvent non contrôlée par la machinerie cellulaire bouleverse cette dernière. Parfois, deux voire trois des effets possibles ci-dessus coexistent et se conjuguent dans une même cellule.

### **1.1.2.2 Les différents types de mutagenèse insertionnelle**

Comme on l'a vu précédemment, les rétrovirus non transformants ne véhiculent pas d'oncogène, mais ils peuvent cependant provoquer une leucémie ou un cancer après un temps de latence de quelques mois. Tout le potentiel cancérigène de ces virus résulte de l'introduction de séquences rétrovirales au sein ou bien à proximité d'un gène cellulaire, dont elles modifient l'expression ou la structure.

Différents mécanismes d'activation sont mis en jeu (Jonkers *et al.* 1996):

#### **1.1.2.2.1 L'insertion de promoteur**

Le rétrovirus est intégré en 5' du gène cellulaire et dans la même orientation (Fig. 1.3). Le promoteur cellulaire est remplacé par le promoteur viral. Les LTRs étant des promoteurs forts et constitutifs, le gène devient surexprimé et échappe aux mécanismes de contrôle cellulaires. Le transcrit est hybride car sa séquence débute dans le LTR.

#### **1.1.2.2.2 Activation transcriptionnelle**

Dans ce cas, le promoteur cellulaire fonctionne et la taille du transcrit est inchangée (Fig. 1.3). Les séquences activatrices des LTRs augmentent le niveau d'expression du gène. Le rétrovirus peut être intégré soit en 5' du promoteur, soit en 3' du gène, soit au milieu de ce gène au sein d'un intron, dans une orientation sens ou anti-sens.

#### **1.1.2.2.3 Élimination de séquences régulatrices du gène cellulaire**

Le rétrovirus est intégré dans la région 3' non traduite du gène cellulaire (Fig. 1.3). Le niveau de transcription est augmenté dû aux séquences activatrices du LTR, mais aussi en raison de la disparition de séquences régulatrices négatives en 3' du gène cellulaire. Un tel mécanisme est retrouvé dans quelques lymphomes T de la souris, comme c'est par exemple le cas pour Notch-1 (Hoemann *et al.* 2000).

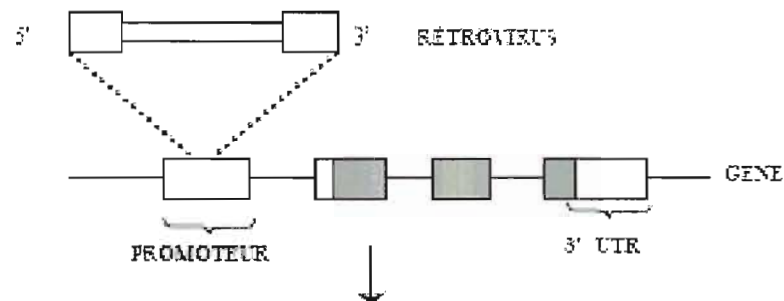
#### **1.1.2.2.4 Interaction longue distance**

Certains sites d'intégration rétrovirale comme le site *pvt-1* décrit dans les lymphomes T de la souris n'ont pas été associés à des gènes cellulaires situés au voisinage de ces sites d'intégration, en dépit de fréquences élevées de réarrangement de cette région dans les leucémies. En fait, l'intégration au locus *pvt-1/Mlvi-1/Mis-1* exerce son effet en augmentant l'expression de *c-myc*, situé à 270 kbp (Lazo *et al.* 1990) (Fig. 1.3).

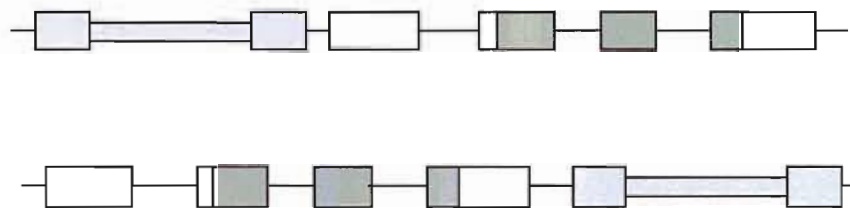
#### **1.1.2.2.5 Modification qualitative de la structure de la protéine**

Ce fait est beaucoup plus rare que les précédents et correspond à une modification de la séquence du proto-oncogène lors de l'intégration virale (Fig. 1.3). Quelques cas sont connus, comme celui de l'érythroblastose aviaire tardive provoquée par le rétrovirus ALV, où l'intégration se produit au sein du gène cellulaire *c-erb B* en éliminant l'extrémité NH<sub>2</sub> terminale de la protéine (Merlino *et al.* 1984).

### Insertion de promoteur



### Activation transcriptionnelle



### Suppression de séquences régulatrices



### Production de protéines chimères

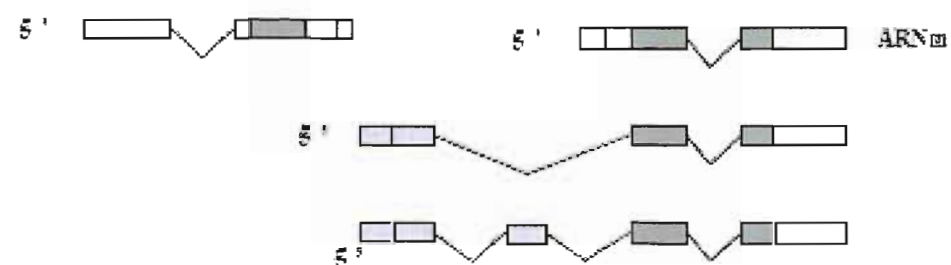


Figure 1.3 Les différents types du mutagenèse insertionnelles (voir pages 6 et 7)

### 1.1.3 LES RÉTROVIRUS EN TANT QU'OUTILS MOLÉCULAIRES POUR LA RECHERCHE D'ONCOGÈNES.

#### 1.1.3.1 Rétrovirus et cancérogenèse : historique

Un peu d'historique va permettre de comprendre à quel point l'histoire de la cancérologie moderne est indissociable de celle des rétrovirus. En effet, c'est avec la découverte des rétrovirus que les premières expériences sur le caractère transmissible des tumeurs ont été faites. C'est également à partir de l'étude du génome des rétrovirus que plus récemment les oncogènes ont été découverts.

L'histoire débute en 1908 quand le premier virus responsable de l'apparition de tumeur fut isolé par deux pathologistes Danois, V. Ellerman et O. Bang (Ellerman *et al.* 1908). Ils décrivirent la possibilité de transmettre une leucémie érythroblastique de poulet à partir d'un broyat cellulaire filtré, mais aucun crédit sérieux n'est accordé à cette nouvelle découverte. L'histoire des virus oncogènes ne débute vraiment qu'en 1910 avec Peyton Rous (Rous 1910 et 1911). Ce dernier étudiait le caractère transmissible des tumeurs chez le poulet. Lorsqu'il transplantait une tumeur (sarcome) de poulet chez un autre poulet, ce dernier développait plus rapidement une tumeur de masse plus importante. Rous entreprit par la suite d'étudier le caractère transmissible de la tumeur à partir d'un filtrat acellulaire. La surprise fut de taille lorsqu'il se rendit compte que ce filtrat était capable de transmettre la tumeur, et ceci en un temps très bref. Il venait de mettre en évidence l'existence d'un agent filtrable, en d'autre terme un virus, contenant l'information nécessaire au développement d'une tumeur.

Puis Ludwig Gross (Gross 1951) ouvrit l'ère moderne de la cancérologie virale en découvrant que certaines leucémies de souris pouvaient être transmises par des virus à condition d'injecter ces virus à des souriceaux nouveaux nés. En 1955, avec le développement de la microscopie électronique, les chercheurs purent voir les particules virales en question (Selby 1954), et de nombreux virus murins responsables de tumeurs chez la souris, comme le virus de Friend ou de Moloney, furent isolés (Friend 1957, Moloney 1960). Certaines leucémies induites par ces virus se développaient avec un temps de latence de quelques mois (dans le cas de rétrovirus non transformants), alors que d'autres se comportaient comme des tumeurs à développement immédiat (rétrovirus transformants). Le premier rétrovirus responsable de l'apparition de tumeurs solides chez la souris fut isolé par Jennifer Harvey en

1964 (Harvey 1964).

La structure génétique de ces rétrovirus a été abordée dans les années 70, avec le développement des techniques de génétique moléculaire. Il a été mis en évidence que les rétrovirus transformants portaient l'information génétique nécessaire au développement tumoral à travers les oncogènes viraux. Peu de temps après, il fut démontré que ces oncogènes viraux étaient en fait d'origine cellulaire (Stéhelin *et al.* 1976).

### **1.1.3.2 Rétrovirus non transformants et cancérogenèse : les sites communs d'intégration.**

#### **1.1.3.2.1 La notion de site commun d'intégration comme hypothèse de travail**

La capacité des rétrovirus non transformants à induire des tumeurs vient du fait qu'ils s'intègrent dans l'ADN cellulaire, ce qui provoque la dérégulation de l'expression des gènes adjacents. C'est cette propriété des rétrovirus à s'intégrer dans le génome cellulaire qui a fait d'eux des outils pour l'identification d'oncogènes. En effet, comme les séquences virales sont connues, le virus intégré constitue une véritable étiquette moléculaire. C'est ce que l'on appelle l'étiquetage rétroviral (*retroviral tagging* en anglais). Ceci nous permet donc de repérer les sites d'intégration rétrovirale dans l'ADN cellulaire, et de connaître les gènes adjacents éventuellement ciblés par le rétrovirus. Lors d'une intégration rétrovirale au niveau d'un locus cellulaire, le profil de restriction de ce dernier est modifié, et l'on nomme ces modifications réarrangements. Les séquences cellulaires adjacentes au provirus sont alors utilisées comme sonde pour trouver des réarrangements de l'ADN cellulaire au niveau d'un ensemble de tumeurs induites par le rétrovirus. Si une sonde détecte un réarrangement dans des tumeurs d'origine distincte, ce locus cellulaire d'intégration virale est qualifié de site commun d'intégration (SCI). Les probabilités pour qu'une intégration virale ait lieu au même endroit étant infimes, si cela se produit, c'est que cet événement procure un avantage sélectif de croissance à la cellule, aboutissant à la multiplication clonale de cette cellule et à la formation de tumeur. En conséquence, il est admis qu'un SCI est significativement associé à la tumorigenèse. Cette hypothèse constitue la base de la recherche d'oncogènes à l'aide de rétrovirus.



### 1.1.3.2.2 Recherche de SCI et découverte d'oncogènes par étiquetage rétroviral

La première activation d'oncogène par un rétrovirus est rapportée en 1981, avec l'activation de *c-myc* par le rétrovirus ALV (*Avian Leukosis Virus*) (Hayward *et al.* 1981). Par la suite, l'application de l'hypothèse des SCI à la recherche d'oncogènes débute avec la découverte de *Pim-1*. Ce SCI relatif au rétrovirus Moloney (MoMuLV) est reporté pour être impliqué dans 50% des lymphomes T analysés (Cuypers *et al.* 1984), avec activation du gène *Pim-1* associé à ce locus (Selten *et al.* 1985). De nombreux autres oncogènes impliqués dans différentes leucémies seront découverts de la même façon, grâce à l'utilisation de plusieurs types de rétrovirus (Jonkers *et al.* 1996).

Parallèlement, l'étiquetage rétroviral a permis des études très élégantes sur la collaboration entre oncogènes dans le développement tumoral. Le principe consistait en l'infection rétrovirale d'une souris transgénique surexprimant un oncogène donné. L'accès aux SCI dans ces souris permettait de découvrir les gènes collaborateurs de l'oncogène qu'elles surexprimaient. Par exemple, 80% des souris *Eμ-Pim-1* (surexprimant le gène *Pim-1*) arboraient une intégration rétrovirale au niveau de *c-myc*, révélant ainsi une collaboration évidente entre ces deux oncogènes (van Lohuizen *et al.* 1989). Réciproquement, des souris *Eμ-myc* possédaient une intégration au niveau de *Pim-1* dans 35% des cas (van Lohuizen *et al.* 1991). Deux autres oncogènes, *Bmi-1* et *Pal-1*, sont également ciblés conjointement avec *Pim-1* dans les souris *Eμ-myc*. Mais de façon intéressante, *Bmi-1* et *Pal-1* ne sont jamais ciblés en même temps, c'est l'un ou l'autre. Ces observations ont permis d'introduire la notion importante de groupes de complémentation. En effet, *Bmi-1* et *Pal-1* ne sont jamais conjointement activés, car en termes de tumorigenèse ils appartiennent au même groupe de complémentation : leurs actions seraient en quelques sortes redondantes. Par contre, *c-myc*, *Pim-1* et *Bmi-1/Pal-1* appartiennent à des groupes de complémentation différents, et sont donc des collaborateurs. Dans la même logique et dans le but de trouver des gènes appartenant à un même groupe, la mutation de *Pim-1* dans des souris *Eμ-myc* a permis de découvrir l'activation rétrovirale d'un nouveau gène dit «compensatoire» de *Pim-1*, appelé *Pim-2* (van der Lugt *et al.* 1995).

Par la suite, la publication du séquençage du génome de la souris en 2002 a permis de révéler tout le potentiel de l'étiquetage rétroviral dans la recherche d'oncogène. Avec l'accès aux séquences, chaque intégration rétrovirale peut être dès lors positionnée précisément dans le génome, permettant ainsi une analyse détaillée et rapide du contenu génétique du locus

d'intégration. Les rétrovirus apparaissent alors comme les puissants outils de l'étiquetage rétroviral à haut débit (*high-throughput retroviral tagging* en anglais). Grâce à la mise au point de techniques de repérage rapide d'intégration rétrovirale basées sur la PCR (Li *et al.* 1999, Sørensen *et al.* 1996, Mikkers *et al.* 2002), des centaines de SCI relatifs à différents types de tumeurs ont été identifiés (Mikkers *et al.* 2002, Hwang *et al.* 2002, Joosten *et al.* 2002, Suzuki *et al.* 2002) (voir Tableau 1.1). L'ensemble de ces données sont archivées dans une banque de données informatique publiquement accessible (<http://RTCGD.ncifcrf.gov>, Akagi *et al.* 2004). Une représentation schématique de la technique PCR que nous avons utilisée pour la recherche de SCI est donnée (Fig. 1.4).

#### 1.1.3.2.3 Généralités sur le rétrovirus RadLV (*Radiation Leukemia Virus*)

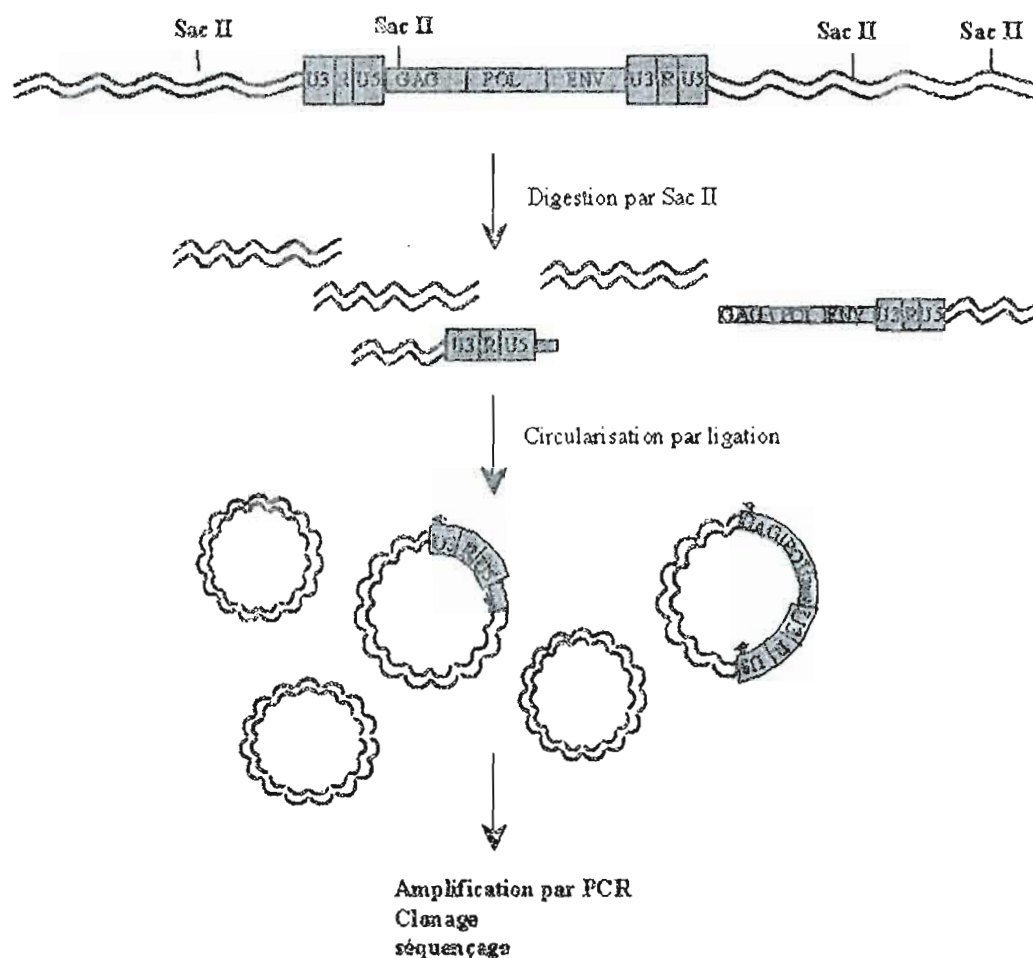
Le RadLV est un rétrovirus endogène de la souris, que nous utilisons comme outil dans notre laboratoire. Ce type de rétrovirus fait partie intégrante du génome de leur hôte, et ils sont transmis selon un mode vertical. Les rétrovirus endogènes murins ne sont pas pathogènes pour leur hôte. Ils peuvent le devenir lorsqu'une recombinaison survient entre un virus endogène écotrope (c'est-à-dire un virus qui n'infecte que des cellules de souris) et un virus non écotrope, aboutissant à la formation d'un virus recombinant éventuellement pathogène. De plus, selon la lignée de souris, ces virus s'expriment plus ou moins tôt dans la vie de l'animal, ou ne s'expriment jamais. Dans le cas de RadLV, l'expression virale peut-être provoquée par irradiation. C'est de cette façon qu'il a d'ailleurs été découvert : lorsque des souris étaient irradiées, ces dernières développaient des thymomes avec un temps de latence de 4 à 6 mois. Par la suite, RadLV a été associé au développement de ces tumeurs (Decleve *et al.*, 1978, Lieberman *et al.* 1979).

Nous travaillons avec un rétrovirus issu d'une lignée cellulaire lymphoïde (lignée BL/VL3) (Rassart *et al.* 1983) produisant des RadLV qui ont les caractéristiques d'être hautement leucémogènes et thymotropes. Plusieurs de ces virus ont été clonés, et l'un de ces clones

SCI	Gènes candidats	Fonction	Chromosome
<i>Dkrni1</i>	<i>Dst</i>	Actin cross-linking protein	6p11-p12
<i>Dkrni2</i>	<i>Ly108</i>	Carcinoembryonic antigen	1
<i>Cis1</i>	<i>Ptma</i>	nuclear protein	2q35-q34
<i>Nkis3</i>	<i>Zfx1b</i>	zing-finger homeobox protein	2q22
<i>Dkrni3</i>	<i>Ptpn1</i>	TYR phosphatase	20q13.1-13.2
<i>Dkrni4</i>	<i>Set/ ND/</i> 1190004A01Rik	nucleosome assembly protein/ protein kinase/ ND	9q34
<i>Gfib</i>	<i>Gfi1b</i>	transcription factor	9q34.13
<i>Notch1</i>	<i>Notch1</i>	receptor	9q34.3
<i>Bmi1</i>	<i>Bmi1</i>	polycomb protein	10p13
<i>Evi18</i>	<i>RasGrp1</i>	RAS exchange factor	15q15
<i>Dkrni5</i>	ND	ND	20
<i>Dkrni6</i>	<i>Pkig</i>	protein kinase inhibitor	20q12-q13.1
<i>Dkrni7</i>	<i>Mcl1</i>	BCL-2 related	1q21
<i>Dkrni8</i>	<i>Cla3</i>	ND	1q21.1
<i>Lef1</i>	<i>Lef1</i>	transcription factor	4q23-q25
<i>Evi55</i>	<i>Camk2d</i>	SER/THR kinase	4
<i>Cis2</i>	ND	ND	9
<i>Nki11</i>	<i>Runx3</i>	transcription factor	1p36
<i>Evi62</i>	<i>E2f2/Idb3</i>	HLH factor/ transcription factor	1p36
<i>Evi143</i>	<i>Ak4/ Lepr</i>	adenylate kinase/ leptine receptor	9p24-p13
<i>Evi58</i>	5830400A04Rik	RAS related	4p13
<i>Gfi1/Evi5</i>	<i>Gfi1/ Evi5</i>	transcription repressor/ cell cycle protein	1p22
<i>Dkrni9</i>	<i>Kdr</i>	TYR kinase receptor	4q12
<i>Kit</i>	<i>Kit</i>	TYR kinase receptor	4q12
<i>Dkrni10</i>	ND	heparanase	4q21.3
<i>Nki16</i>	ND	SER/THR kinase	12q24.31
<i>Evi65</i>	<i>Coro1c/ Selp1</i>	actin binding protein/ selectin	12q24.1
<i>Evi78</i>	<i>Caln2/ ND</i>	calcium binding protein/ ribosomal protein	2p21
<i>Ccnd2</i>	<i>Ccnd2</i>	cell cycle regulator	12p13
<i>Cis3</i>	ND/ <i>Wnt5b</i>	F-box protein/ growth factor	12p13.3
<i>Evi167</i>	<i>Sema4b</i>	receptor	15q26.1
<i>Cis4</i>	ND	ND	ND
<i>Dkrni11</i>	ND	ND	10q25

**Tableau 1.1 Exemple de recherche d'oncogène obtenue par étiquetage rétroviral à haut débit.** Le rétrovirus Moloney (MoMuLV) utilisé dans cette étude induit des leucémies de type T (Mikkers *et al.* 2002). ND, non-déterminé.

(clone V13) (Rassart *et al.* 1986) possédait un des potentiels leucémogènes les plus élevés chez les rétrovirus non transformants. Ce clone est utilisé dans notre laboratoire pour la recherche de SCI par l'étiquetage rétroviral.



**Figure 1.4 Technique de PCR inversée.** Cette technique permet l'amplification des séquences d'ADN cellulaire inconnues qui sont adjacentes au provirus (Li *et al.* 1999). L'ADN génomique d'une tumeur est digéré par l'enzyme *Sac* II. Cette enzyme a été choisie pour son site de restriction riche en GC, plus fréquent dans les îlots CpG au niveau des promoteurs des gènes. Les produits de digestion sont ensuite circularisés sur eux mêmes par ligation. À l'aide d'oligonucléotides situés de part et d'autre du fragment proviral (usuellement dans le LTR et dans GAG), les séquences d'ADN cellulaire adjacentes sont amplifiées par PCR, clonées puis séquencées.

## 1.2 ÉTAT DES CONNAISSANCES GÉNÉRALES SUR LES MICROARN

### 1.2.1 BIOGENÈSE DES MICROARN

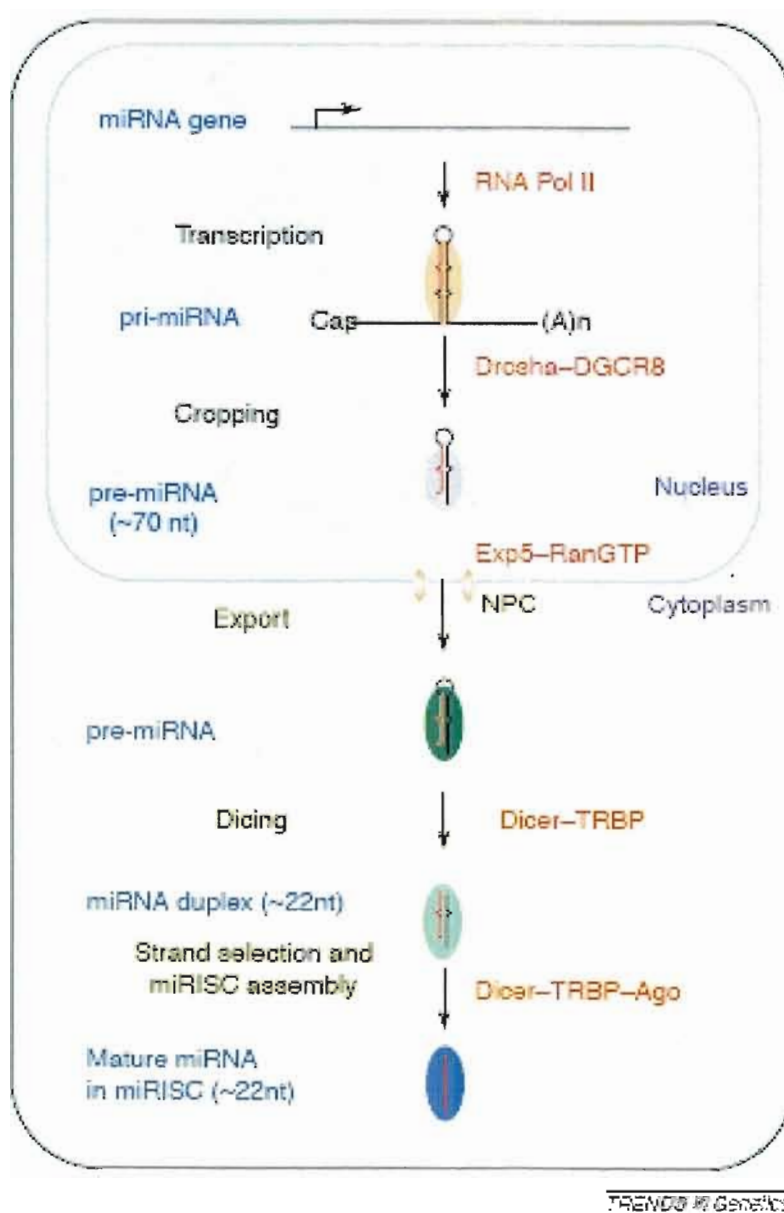
Les miARN apparaissent pour la première fois dans la littérature en 1993 lorsque Lee *et al.* (1993) rapportent chez *Caenorhabditis elegans* (*C. elegans*) l'existence de *lin-4*. Ce petit ARN de 21pb avait la particularité d'être complémentaire à sept séquences, précédemment impliquées comme éléments négatifs de régulation (Wightman *et al.* 1991), présentes dans le 3'UTR de l'ARNm de *lin-14*. L'interaction de *lin-4* avec ces séquences cibles provoquait une inhibition d'expression post-transcriptionnelle de *lin-14* (Wightman *et al.* 1993). Sept ans plus tard, des dizaines de miARN seront clonés et officiellement reconnus comme une nouvelle classe d'éléments régulateurs post-transcriptionnels, hautement conservés entre espèces (Lagos-Quintana *et al.* 2001, Lau *et al.* 2001, Lee *et al.* 2001). Depuis la vague de clonage de miARN de 2001, des centaines de miARN ont été découverts (<http://www.sanger.ac.uk/Software/Rfam/mirna/>, Griffiths-Jones 2004 et 2006, Ambros 2003). Aujourd'hui, 114 miARN sont répertoriés pour *C. elegans*, 78 pour *Drosophila melanogaster*, 369 pour *Danio rerio* (Zebrafish), 249 pour *Mus musculus* (souris), 326 chez l'humain et 117 pour *Arabidopsis thaliana*. Les prédictions annoncent qu'il existerait entre 500 et 1000 miARN par espèce, ce qui représenterait 2 à 3 % de l'ensemble des gènes.

#### 1.2.1.1 Mécanisme de production des microARN

De nombreuses revues résument très bien l'ensemble des mécanismes de la biogenèse des miARN. Les connaissances générales décrites dans ce paragraphe sont issues d'une des revues les plus récentes (Zhang *et al.* 2006), que j'ai agrémenté des toutes dernières informations disponibles.

Les miRNAs sont de petits ARN d'une vingtaine de nucléotides. Ils sont issus du découpage d'un ARN beaucoup plus long, appelé le pri-miRNA, transcrit par l'ARN polymérase II (Fig. 1.5). Le pri-miRNA possède donc un cap en 5', une queue polyA en 3', et peut être épissé. Une fois transcrit, il prend une structure secondaire spécifique contenant des tiges-boucles qui sont reconnues par un énorme complexe protéique (500 à 650 kDa) appelé

le *microprocessor*. Ce dernier comprend essentiellement deux protéines : une RNase III endonucléase appelée Drosha, et le co-facteur DGCR8 (Pasha chez la Drosophile). DGCR8 est une protéine qui possède deux domaines de liaison à l'ARN double brin, et permet donc l'ancrage du microprocessor au niveau des structures tiges-boucles du pri-miRNA. Drosha coupe le pri-miRNA de façon asymétrique et spécifique au niveau de ces structures. Ces petites structures en épingle à cheveux (60 à 110 nt) sont appelées les pre-miRNAs, et sont caractérisées par un groupement phosphate en 5' ainsi que deux nucléotides non-appariés en 3'. Les pre-miRNAs sont ensuite transportés dans le cytoplasme par un transporteur situé dans la membrane nucléaire appelé Exportin-5 (Exp-5, membre de la famille des récepteurs transporteurs RAN). Le transport requiert de l'énergie et la structure spécifique du pre-miRNA décrite plus haut. Une fois dans le cytoplasme, les pre-miARN sont incorporés dans un complexe protéique appelé RISC (*RNA-induced silencing complex*). Le RISC comporte au moins quatre protéines clairement identifiées chez l'humain: Dicer, TRBP, AGO2 et RCK/p54 (Gregory *et al.* 2005, Chu et Rana 2006). Dicer, une autre endonucléase de type RNase III, coupe le pre-miRNA pour donner un petit ARN double brin de 20-23 nt composé du miARN et de sa séquence complémentaire. Les deux brins sont ensuite séparés par une activité hélicase associée au RISC, et le brin correspondant au miARN mature est conservé. Le complexe miARN-RISC s'associe ensuite aux 3'UTR des ARN messagers cibles du miARN pour inhiber leur traduction. Le rôle de TRBP dans le RISC n'est pas établi, mais il a été montré que cette protéine facilitait le découpage du pre-miRNA (Haase *et al.* 2005). AGO2 interagit directement avec le miARN mature, et permettrait l'ancrage spécifique du complexe miARN-RISC sur l'ARN cible (Parker *et al.* 2005). RCK/p54 est directement impliqué dans l'inhibition de la traduction dirigée par le RISC, et séquestre le complexe miARN-RISC-ARNm dans un domaine cellulaire appelé *P-bodies* (Chu et Rana 2006). L'ARNm stocké dans les *P-bodies* pourra être dégradé, ou bien réintégré le circuit de traduction protéique. Si RCK/p54 apparaît nécessaire à la séquestration dans les *P-bodies*, cette localisation n'est cependant pas nécessaire à la fonction du RISC. En effet, lorsque la complémentarité de séquence entre le miARN et sa séquence cible est totale, le mécanisme observé est identique à celui des siRNAs (silencing RNAs) avec dégradation de l'ARNm, sans passage par les *P-bodies*.



**Figure 1.5 Production des miARN.** Le pri-miARN est transcrit par l'ARN polymérase II pour être découpée en pre-miARN par l'enzyme Drosha. Une fois exportés vers le cytoplasme, les pre-miRNA sont découpés par l'enzyme Dicer. Les miRNAs résultants sont incorporés dans le complexe protéique RISC, pour aller se fixer sur le 3'UTR de leur mRNA cible, et inhiber leur traduction (extrait de Kim *et al.* 2006).

### 1.2.1.2 Organisation génomique des microARN

La plupart du temps, les miARN sont issus d'une unité de transcription indépendante de celle de tout autre gène (Lagos-Quintana *et al.* 2001, Lau *et al.* 2001, Lee *et al.* 2001). Ceci dit, une proportion significative de miARN (un quart des miARN chez l'humain) est produite à partir d'un intron d'un autre gène (Aravin *et al.* 2003, Lagos-Quintana *et al.* 2003, Lai *et al.* 2003, Lim *et al.* 2003a). Ces gènes peuvent être codants (80% des cas) ou non codants (20%). Le plus souvent, les miARN introniques sont dans la même orientation que celle du gène hôte, suggérant que la production de ces miARN dépend du promoteur de ce gène. D'autres miARN sont regroupés dans le génome et produits à partir d'un seul et même messenger (groupe de miARN). Chez l'humain et *C. elegans*, la majorité des miARN sont isolés (Lim *et al.* 2003a et b), alors que plus de la moitié des miARN chez la Drosophile sont groupés (Aravin *et al.* 2003). Les miARN au sein d'un groupe sont souvent relativement homologues, mais pas toujours (Lagos-Quintana *et al.* 2001, Lau *et al.* 2001). Par exemple, les miARN orthologues de *lin-4* et *let-7* chez la Drosophile et l'humain forment un groupé, et sont co-exprimés à partir d'un même transcrit (Aravin *et al.* 2003, Bashirullah *et al.* 2003, Sempere *et al.* 2003). De façon intéressante, en dépit de leur séquence différente, ces deux miARN sont impliqués dans le développement. Ceci montre que même si les miARN d'un groupe n'ont aucune homologie, leurs fonctions peuvent être reliées.

L'hétérogénéité qui règne dans l'organisation génomique des miARN constitue un obstacle à leur identification, et serait à l'origine d'une sous-estimation de leur nombre (Bartel *et al.* 2005). Ainsi, certains miARN n'ont probablement pas été identifiés car difficiles à cloner, et/ou difficile à débusquer informatiquement. Jusqu'à présent, les prédictions ont été établies selon des critères bien précis (Ambros *et al.* 2003). Premièrement, le miARN doit être exprimé comme un ARN distinct de 21 nucléotides clairement identifiable, par Northern-blot ou autre technique expérimentale. Deuxièmement, le miARN doit provenir d'une structure tige-boucle assez caractéristique (Fig. 1.6): usuellement, ces structures s'étendent sur 60 à 110 nucléotides, et le miARN est contenu dans la partie tige. Troisièmement, le miARN mature devrait être généré par l'enzyme Dicer. Enfin, la conservation entre différentes espèces est prise en compte. Cependant, il est fort probable que



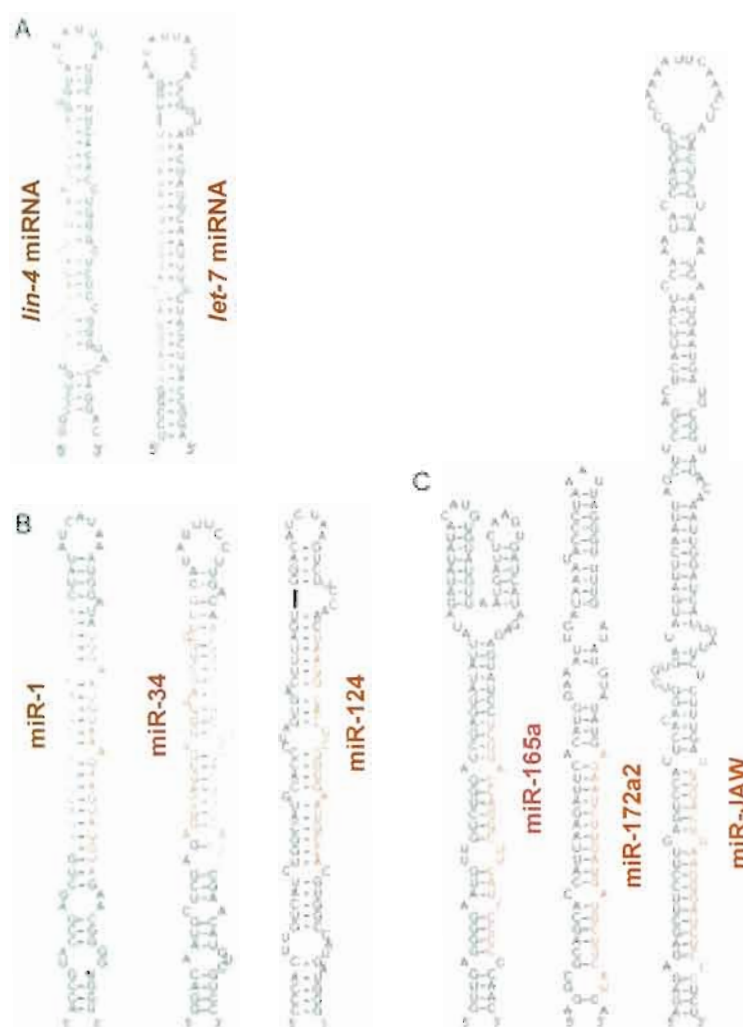
tous les miARN ne soient pas conservés (bien que se soit majoritairement le cas) et que chaque espèce ait des miARN qui leur soient propres.

### 1.2.1.3 Profil d'expression des microARN

De nombreuses études par micropuces ont révélé un profil d'expression complexe de l'ensemble des miARN, qui varie selon les tissus et leurs stades de développement. Il apparaît donc que chaque cellule possède un profil d'expression de miARN distinct qui oeuvre dans la régulation post-transcriptionnelle de son transcriptome, et ceci pour chaque étape de son développement ou en réponse aux stimuli externes (Bartel *et al.* 2004). De plus, les quantités de chaque miARN dans une cellule peuvent varier énormément, certains miARN étant plus abondants que d'autres.

### 1.2.1.4 Régulation de l'expression des microARN

La régulation de l'expression des miARN est un point intéressant qui commence tout juste à être investigué. Tout d'abord, étant donné que les gènes qui produisent les miARN sont transcrits par l'ARN polymérase II, tous les éléments de régulation «classiques» qui régissent ce système de transcription s'appliquent. Il s'agit notamment des régulations qui concernent les promoteurs de ces gènes, ou des régulations post-transcriptionnelles régies par les 5' et 3' UTRs des ARN transcrits. Mais à l'image des différents types d'organisations génomiques des miARN, il semblerait qu'il existe une dualité dans leur système de transcription. Récemment, il a été publié que 20% des miARN chez l'humain résident dans des éléments de séquence répétitifs (Borchert *et al.* 2006). Les auteurs montrent que des miARN situés dans une région du chromosome 19 humain, riche en séquences ALU, sont transcrits par l'ARN polymérase III. Cette étude établit donc un lien entre les éléments transposables et l'expression des miARN.



**Figure 1.6 Exemples de structures tige-boucle des pre-miARN (extrait de Bartel *et al.* 2004).** (A) Exemple des pre-miARN de *lin-4* et *let-7* chez *C. elegans*, dont des homologues existent chez la Drosophile et les mammifères. (B) Pre-miARN de *miR-1*, *miR-34* et *miR-124* de *C. elegans*, qui ont de proches homologues chez la Drosophile et les mammifères. (C) Pre-miARN de *miR-165a*, *miR-172a2* et *miR-JAW* chez *Arabidopsis Thaliana*. Notez la longueur inhabituelle du pre-miARN de *miR-JAW*.

À priori, n'importe quelle étape du système de production des miARN peut être soumise à une régulation qui lui est propre (Fig. 1.5). Des évidences concernant une régulation au niveau du découpage du pre-miARN et du pri-miARN ont été rapportées. Obernosterer *et al.* (2006) décrivent en effet que miR-138 n'est détectable que dans le cerveau et le foie fœtal, alors que son pre-miARN est présent dans plusieurs types de tissus. Comme le pre-miARN est détecté, cela suggère fortement une régulation au niveau de Dicer. Par la suite, Thomson *et al.* (2006) ont rapporté un niveau constant du pri-miARN de let-7g au cours du développement embryonnaire chez la souris, alors que le miARN mature n'est détectable qu'au 10<sup>ème</sup> jour post-fécondation. Ils ont voulu savoir si cette observation s'appliquait à d'autres miARN, et sur 12 miARN différents analysés, 9 ont présenté des niveaux d'expression différents entre pri-miARN et miARN matures. Ceci met en évidence une régulation au niveau du processing du pri-miARN par Drosha. Pour corroborer ces résultats, des différences d'expression au sein des miARN d'un groupe sont souvent notables dans les analyses d'expression de miARN faites par micropuces, suggérant une régulation au niveau du processing par Drosha et/ou Dicer. Pour l'instant, rien n'est connu quant au(x) mécanisme(s) mis en jeu. Une étude récente apporte cependant une nouvelle pièce au puzzle en démontrant que DGCR8 lie l'hème (Faller *et al.* 2006). Cette liaison permet la dimérisation de DGCR8, et la fixation subséquente au pri-miARN de trois de ces dimères actifs dans le découpage. En l'absence d'hème, DGCR8 reste sous forme monomérique inactive. Les auteurs proposent que la multimérisation de DGCR8, contrôlée par l'hème, soit un déterminant de la régulation post-transcriptionnelle de l'expression des miARN.

## 1.2.2 FONCTIONS DES MICROARN

### 1.2.2.1 Inhibition de l'expression des gènes cibles et implication dans de nombreux processus biologiques.

La fonction communément admise des miARN consiste en la régulation négative de l'expression de leurs gènes cibles, principalement par inhibition de la traduction. Ce fait a été établi de façon évidente pour plusieurs miARN, comme *lin-4* inhibant l'expression de *lin-14* et *lin-28* par exemple (Lee *et al.* 1993, Wightman *et al.* 1993). Le blocage semble intervenir

après l'initiation de la traduction (Olsen *et al.* 1999, Seggerson *et al.* 2002). La principale observation de ces études relate une diminution de l'expression de la protéine, sans variation notable du niveau du transcrit. Cependant, il a été observé que l'inhibition de l'expression d'un gène par un miARN peut se faire par dégradation de l'ARN cible (Yekta *et al.* 2004). Cette disparité dans le mécanisme d'inhibition relatif aux miARN a été corrélée à deux types différents d'interactions miARN-ARNm : lorsque la complémentarité de séquence entre un miARN et sa séquence cible est parfaite, l'ARNm serait dégradé, rejoignant alors un mécanisme de type siARN (Yekta *et al.* 2004). C'est d'ailleurs le cas chez les plantes (Schwab *et al.* 2005). Dans le cas d'une complémentarité imparfaite entre un miARN et sa séquence cible, la traduction serait inhibée sans que le niveau de l'ARNm ne soit affecté. Ceci est majoritairement observé dans le règne animal, où l'identité parfaite entre un miARN et sa séquence cible n'est souvent observée que dans la portion 5' du miARN (nucléotide 2 à 8 appelés «5' seed», Brennecke *et al.* 2005). Cette catégorisation du mécanisme d'inhibition des miARN est à présent remise en question. Il semblerait que les miARN puissent diminuer le niveau d'un transcrit de façon plus commune que précédemment admis, et ce même lorsque les séquences du miARN et de sa cible sont imparfaitement complémentaires (Lim *et al.* 2005, Bagga *et al.* 2005).

Étant donné le nombre de miARN et le nombre de cibles qui sont attribuées à chacun d'eux, il apparaît que les miARN participent à la régulation de nombreux processus biologiques incluant par exemple l'apoptose, la prolifération, le développement, ou la différenciation cellulaire (Ambros *et al.* 2004, Alvarez-Garcia *et al.* 2005).

#### 1.2.2.1.1 Recherche des gènes cibles des microARN

Un des défis majeurs dans le domaine des miARN concerne l'identification de leurs gènes cibles. Pour cela, de nombreuses prédictions informatiques sont disponibles et mises à disposition de tous par internet (Tableau 1.2). Le problème des prédictions consiste à éviter les faux positifs, tout en ne passant pas à côté de véritables sites régulateurs des miARN. Un des moyens utilisé dans les prédictions est l'alignement inter-espèces des 3'UTR des gènes : les sites de miARN conservés entre espèces ont plus de chance d'avoir une signification biologique. Certains principes d'hybridation des miARN à leurs sites cibles dans les 3'UTR

ont également été établis. Ainsi de façon unanime, il semblerait que l'extrémité 5' des miARN soit particulièrement importante à leur fonction inhibitrice. Cette partie, appelée le 5' seed, comprend les 8 premiers nucléotides ou les nucléotides 2 à 8 de l'extrémité 5' (Brennecke *et al.* 2005, Doench *et al.* 2004). Le plus souvent, une identité parfaite entre le 5' seed et la séquence cible est requise pour que l'inhibition ait lieu. Dans ce cas, l'extrémité 3' du miARN n'apparaît pas capitale pour sa fonction. Néanmoins, elle pourrait avoir un rôle dans la spécificité des miARN au sein d'une même famille (qui possède donc le même 5' seed). Dans les cas où l'identité entre 5' seed et séquence cible est faible, l'extrémité 3' peut avoir un rôle compensatoire, et l'identité du miARN avec sa cible au niveau de cette région est alors plus forte qu'à l'accoutumé (Brennecke *et al.* 2005) (Fig. 1.7). Il a été également mis en évidence que les wobbles de type G:U situé dans l'extrémité 5' du miARN diminuaient fortement l'effet inhibiteur.

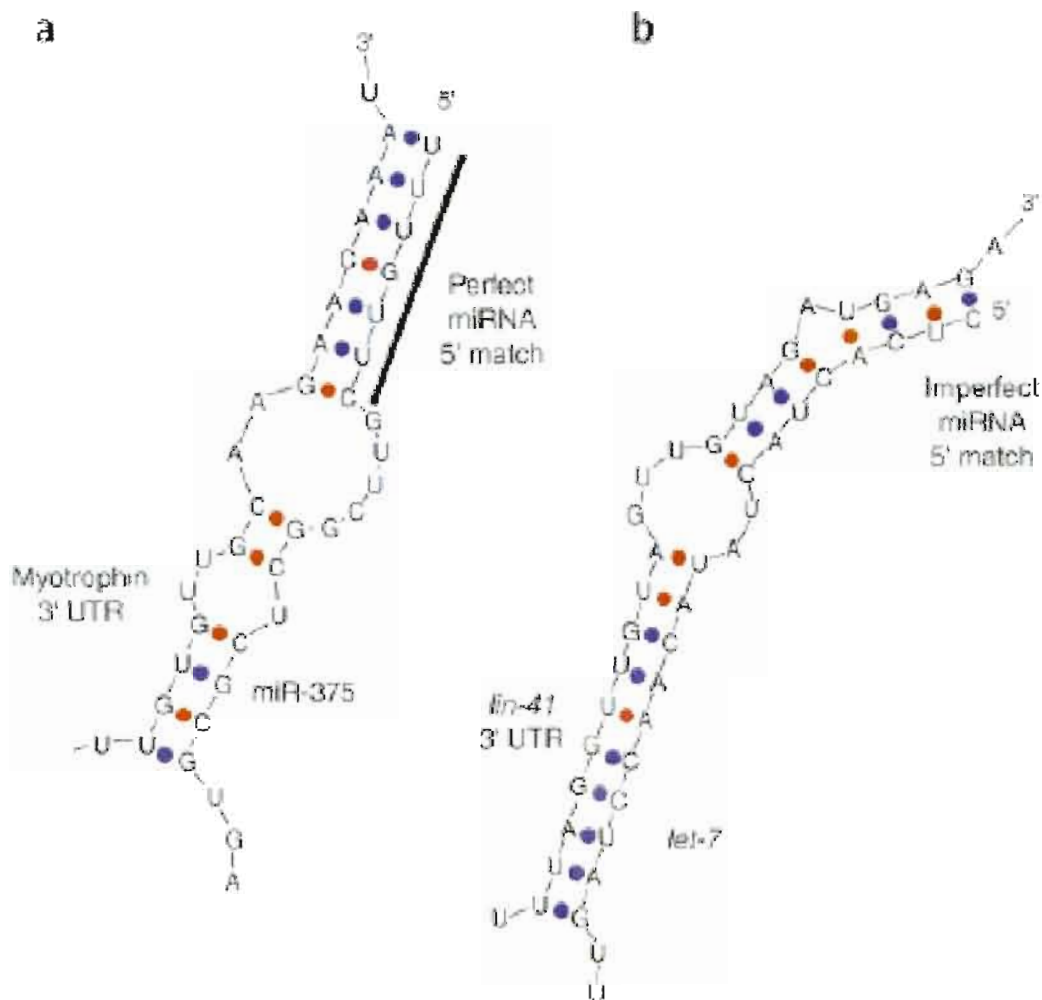
La plupart des prédictions informatiques prennent en compte ces principes, et constituent un précieux outil de travail. De plus, l'existence de plusieurs sites internet de prédiction permet une comparaison des résultats entre eux pour une sélection de gènes cibles plus significative. Ceci dit, l'ensemble des résultats obtenus nécessiteront toujours une confirmation expérimentale *in vivo* pour être validés.

#### 1.2.2.1.2 Subtilités d'inhibition de l'expression des gènes par les miARN

Il semblerait que l'inhibition de l'expression des gènes par les miARN ne soit pas aussi simple qu'il y paraît. Tout d'abord, bon nombre d'ARNm possédant des séquences cibles de miARN dans leur 3'UTR sont activement traduits dans la cellule, et ce même en présence des miARN (Farh *et al.* 2005, Lim *et al.* 2005, Lewis *et al.* 2003). D'ailleurs, il a été montré dans les cellules HeLa que les miARN localisent aux niveaux des polysomes en association avec leurs ARN cibles, ces derniers étant même activement traduits (Maroney *et al.* 2006). Cette étude concorde aussi avec les résultats obtenus chez *C. elegans* cités dans le paragraphe précédent (Olsen *et al.* 1999, Seggerson *et al.* 2002). Les mécanismes de régulation impliquant les miARN semblent donc plus complexes qu'il n'y paraît. Les premiers rapports concernant la régulation du développement par les miARN chez *C. elegans*

**Tableau 1.2 Différents outils informatiques pour la prédiction de cibles de miARN**

<b>Méthodes</b>	<b>Organismes</b>	<b>Web sites</b>
<b><u>Prédictions de cibles de miARN</u></b>		
Prédictions à l'EMBL	Drosophile	<a href="http://www.russell.emblheidelberg.de/miRNAs/">http://www.russell.emblheidelberg.de/miRNAs/</a>
miRanda	Drosophile, vertébrés	<a href="http://www.microrna.org//miranda.html">http://www.microrna.org//miranda.html</a>
miRBase	Insects, vertébrés, nématode	<a href="http://microrna.sanger.ac.uk/targets/v2/">http://microrna.sanger.ac.uk/targets/v2/</a>
PicTar	Drosophile, vertébrés, nématodes	<a href="http://pictar.bio.nyu.edu">http://pictar.bio.nyu.edu</a>
TargetScan, TargetScanS	Vertébrés, drosophile, nématode	<a href="http://genes.mit.edu/targetscan">http://genes.mit.edu/targetscan</a>
<b><u>Localisation de séquences cibles des miARN dans les 3'UTR</u></b>		
RNAhybrid		<a href="http://bibiserv.techfak.unibielefeld.de/rnahybrid/welcome.html">http://bibiserv.techfak.unibielefeld.de/rnahybrid/welcome.html</a>
DIANA-MicroT		<a href="http://diana.pcbi.upenn.edu/DIANA-microT">http://diana.pcbi.upenn.edu/DIANA-microT</a>
RNA22		<a href="http://cbcsrv.watson.ibm.com/rna22.html">http://cbcsrv.watson.ibm.com/rna22.html</a>
<b><u>Bases de données sur les cibles confirmées expérimentalement</u></b>		
Tarbase		<a href="http://www.diana.pcbi.upenn.edu/tarbase.html">http://www.diana.pcbi.upenn.edu/tarbase.html</a>
Argonaute		<a href="http://www.ma.uni-heidelberg.de/apps/zmf/argonaute/interface">http://www.ma.uni-heidelberg.de/apps/zmf/argonaute/interface</a>
miRNAMAP		<a href="http://mirnamap.mbc.nctu.edu.tw/">http://mirnamap.mbc.nctu.edu.tw/</a>



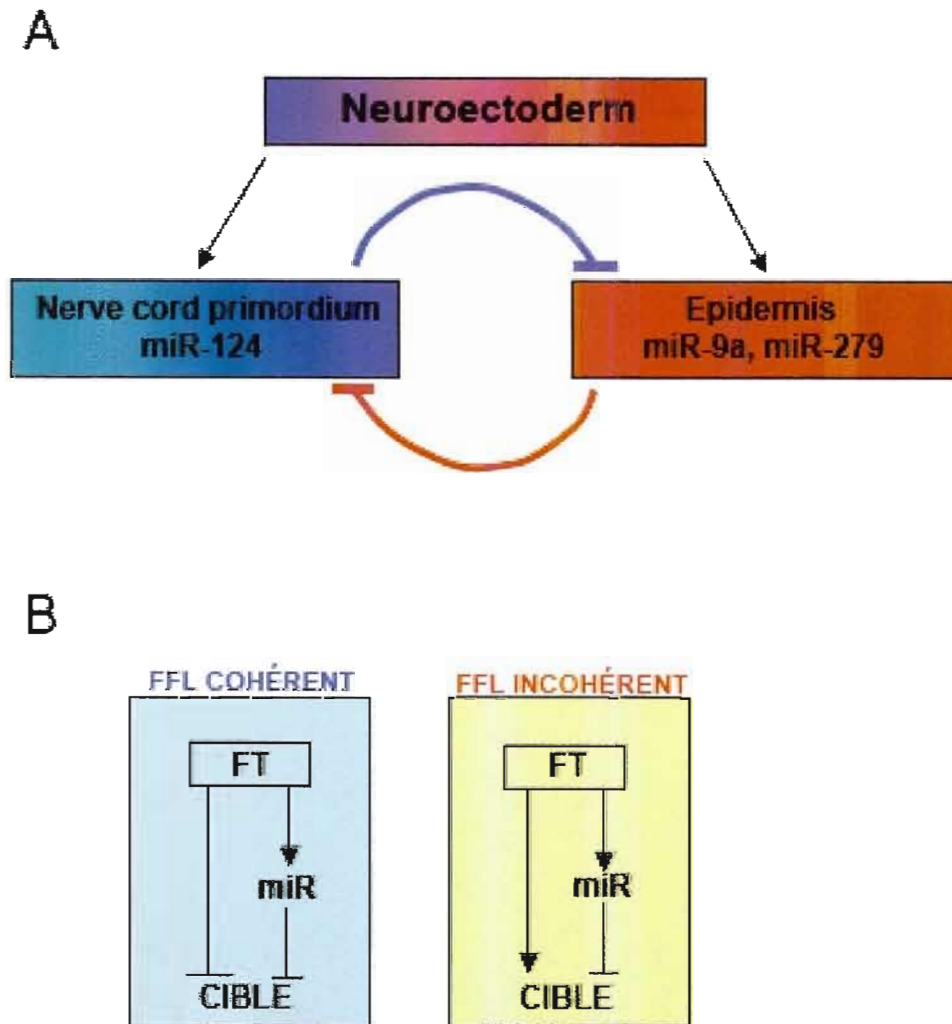
**Figure 1.7 Exemples d'hybridation d'un miARN avec sa séquence cible (extrait de Rajewsky *et al.* 2006).** (A) MiR-375 hybride au 3'UTR du gène de la myotrophine, selon le modèle du 5' seed, où l'identité est parfaite avec la séquence cible. (B) Exemple de *let-7* avec le 3'UTR du gène *lin-41* : l'identité imparfaite du 5'seed avec sa séquence cible est compensée par une forte identité de séquence au niveau de *let-7*.

décrivaient un effet drastique de *lin-4* et *let-7* sur l'expression de leurs gènes cibles (Lee *et al.* 1993, Wightman *et al.* 1993, Reinhart *et al.* 2000, Vella *et al.* 2004). Dans les deux cas, cette forte inhibition peut être corroborée avec la présence de plusieurs sites spécifiques à ces deux miARN au niveau du 3'UTR des gènes ciblés. Or, il apparaît chez l'animal que la plupart des 3'UTR analysés ne contiennent qu'un site spécifique d'un miARN donné, ce qui rend peu probable un fort effet inhibiteur. Par contre, ces 3'UTR contiennent souvent de nombreux sites uniques pour différents miARN, ce qui laisse prévoir des effets subtils à l'échelle d'un seul type de miARN et une collaboration entre différents types de miARN. De plus, mis à part chez la souris où l'inhibition génétique de Dicer est létale chez l'embryon (Bernstein *et al.* 2003), l'effet n'apparaît pas aussi délétère chez les autres vertébrés. Chez *Danio rerio* par exemple, l'absence de Dicer n'empêche pas la différenciation cellulaire et la segmentation embryonnaire, mais crée des problèmes dans la morphogenèse et notamment celle du système nerveux (Giraldez *et al.* 2005). Ainsi de façon surprenante, l'absence de centaines de miARN responsables de l'inhibition de milliers de gènes n'est donc pas aussi spectaculaire qu'attendue. Ceci pousse les chercheurs à penser que, le plus souvent, la fonction première d'un miRNA n'est pas l'inhibition *per se* de ses gènes cibles.

### 1.2.2.2 L'exclusion mutuelle

L'exclusion mutuelle est un modèle de fonctionnement pour les miARN qui a trouvé son fondement dans de troublantes et judicieuses observations. Tout d'abord, l'analyse de 3'UTR chez la *Drosophile* pour la présence de sites de miARN a permis de distinguer deux groupes de gènes: les gènes cibles de miARN et les gènes anti-cibles (Stark *et al.* 2005). Les premiers contiennent des sites pour les miARN dans leur 3'UTR, alors que les deuxièmes n'en contiennent pas. D'une façon générale, les gènes cibles sont impliqués dans le développement, la différenciation cellulaire ou des fonctions particulières, alors que les gènes anti-cibles sont plutôt enrichis de gènes ubiquitaires et impliqués dans des fonctions cellulaires de bases. Ceci est finalement logique, puisque la cellule ne veut pas que ses fonctions de bases soient inhibées et évite donc la présence de sites pour les miARN dans ses «house keeping genes». L'hypothèse d'exclusion mutuelle soutient que les miARN seraient co-exprimés dans la cellule avec des gènes qu'ils ne ciblent pas (les gènes anti-cibles) et non





**Figure 1.8 (A) Exemple du rôle de l'exclusion mutuelle dans le développement (extrait de Hornstein *et al.* 2006).** Chez la *Drosophila*, le neuroectoderme donne naissance à «l'épidermis» et au «nerve cord primordium». MiR-124 est fortement exprimé dans le «nerve cord primordium», alors que ses cibles sont exprimées dans «l'épidermis», et vice versa pour miR-9a et miR-279. Ceci garantirait une parfaite différenciation de chacun des tissus. **(B) Illustration des FFL cohérent et incohérent.** FT, Facteur de Transcription. Dans le modèle cohérent, le FT stimule l'expression du miARN et réprime l'expression de sa cible. Dans le modèle incohérent, le FT stimule l'expression des deux à la fois.

avec les gènes qu'ils ciblent. Plusieurs exemples flagrants et conservés entre espèces existent, comme le miARN miR-1 qui est exclusivement exprimé dans le muscles alors que ses gènes cibles ne sont pas musculaires (Lim *et al.* 2005, Stark *et al.* 2005). MiR-124 est quant à lui exprimé dans les neurones, alors que ses cibles sont ectodermiques. MiR-9 est la réciproque de miR-124, puisqu'il est exprimé dans l'ectoderme alors que ses cibles sont neuronales (Fig. 1.8A). Dans ce cas, les miARN semblent être exprimés pour éliminer toutes traces de transcrits non spécifiques au tissu, et garantir ainsi la robustesse génétique de ce dernier. Pour corroborer le paragraphe précédent à propos des sites uniques de miARN dans les 3'UTR, notons qu'ici les miARN sont exprimés très fortement alors que leurs cibles ne le sont que très faiblement, ou pas du tout. Dans ce cas, la présence d'un seul site pour ces miARN dans le 3'UTR des gènes cibles permet tout de même un effet inhibiteur efficace, qui serait plus subtil voir même indécélable sur une plus grande quantité d'ARN cible. L'exclusion mutuelle résoudrait ainsi le problème de l'inefficacité d'inhibition due à la présence d'un site unique pour un miARN donné dans les 3'UTR des gènes cibles.

Selon ce modèle, il faudrait donc considérer les gènes cibles et les gènes anti-cibles pour déterminer la fonction d'un miARN. Dans le cas de miR-9 par exemple, l'unique étude des gènes cibles aurait amené à conclure qu'il s'agit d'un miARN neuronal, alors qu'il participe plus certainement à l'établissement de l'ectoderme (Fig. 1.8A).

Ceci dit, les miARN n'adhèrent pas tous à la notion d'exclusion mutuelle. Certains miARN ont une expression très variable au sein d'un tissu, et sont exprimés dans des tissus différents. De plus, d'autres miARN forment des familles en raison de leur forte homologie de séquence, et partagent donc les mêmes cibles. Ces miARN sont souvent exprimés à partir de *loci* distincts dans différents types cellulaires, ce qui rend la comparaison de gènes cibles et anti-cibles difficile.

### 1.2.2.3 Les microARN et la canalisation génétique

Selon l'hypothèse évoquée précédemment, les miARN permettraient la stabilisation d'un phénotype en limitant sa variabilité (Hornstein *et al.* 2006). Pour expliquer cela, deux modèles de fonctionnement déjà connus et basés sur le principe «Feed-Forward Loop (FFL)» peuvent être également appliqués aux miARN : le FFL cohérent, et le FFL incohérent (Fig.

1.8B). Dans le FFL cohérent, un même signal (facteur de transcription, par exemple) commande l'expression d'un miARN alors qu'il réprime celle de son gène cible. Le miARN et sa cible sont alors exprimés selon un mode d'exclusion mutuelle où le miARN complémente l'inhibition du promoteur de son gène cible, et assure l'élimination de toute fuite transcriptionnelle. Dans le FFL incohérent, un même signal commande à la fois l'expression du miARN et de sa cible. Ceci peut-être illustré par *c-myc* qui active en même temps l'expression du groupe de miARN *miR-17-92* et celle de E2F1, alors que deux miARN de ce groupe (*miR-17-5p* et *miR-20a*) ciblent E2F1 (O'Donnell *et al.* 2005). Une interprétation de ce phénomène contradictoire serait que, dans ce cas, les miARN serviraient à contrôler rigoureusement la quantité du transcrit E2F1, en «tamponnant» les variations d'activité indésirables du promoteur de E2F1. D'ailleurs, il a été montré récemment que les facteurs E2F activent eux-mêmes l'expression de leur inhibiteur *miR-17-92*, renforçant le rôle de «garde fou» de *miR-20* vis-à-vis de l'expression des facteurs E2F (Sylvestre *et al.* 2006, Wood *et al.* 2006).

Que se soit dans le cas du FFL cohérent, ou dans celui du FFL incohérent, l'expression des miARN permettrait donc d'estomper les variations d'expression de leurs gènes cibles, ce processus étant qualifié de *canalisation génétique*.

### 1.2.3 LES MICROARN ET LE CANCER

Étant donné le caractère multifactoriel du cancer, les miARN sont apparus comme de parfaits effecteurs de la cancérogenèse, puisqu'ils peuvent cibler de nombreux gènes à la fois. Depuis 2002, la littérature ne cesse d'ailleurs de s'enrichir à ce sujet.

#### 1.2.3.1 Évidences de l'implication des microARN dans le cancer

L'ampleur de l'implication des miARN dans le cancer a pu être mise en évidence grâce à de puissants outils tels que les micropuces (Liu 2004, Lu 2005, Nelson 2004 et Calin 2006 pour une revue). Toutes ces analyses ont unanimement révélé une différence significative dans le profil d'expression des miARN entre tissus tumoraux et tissus sains. Les

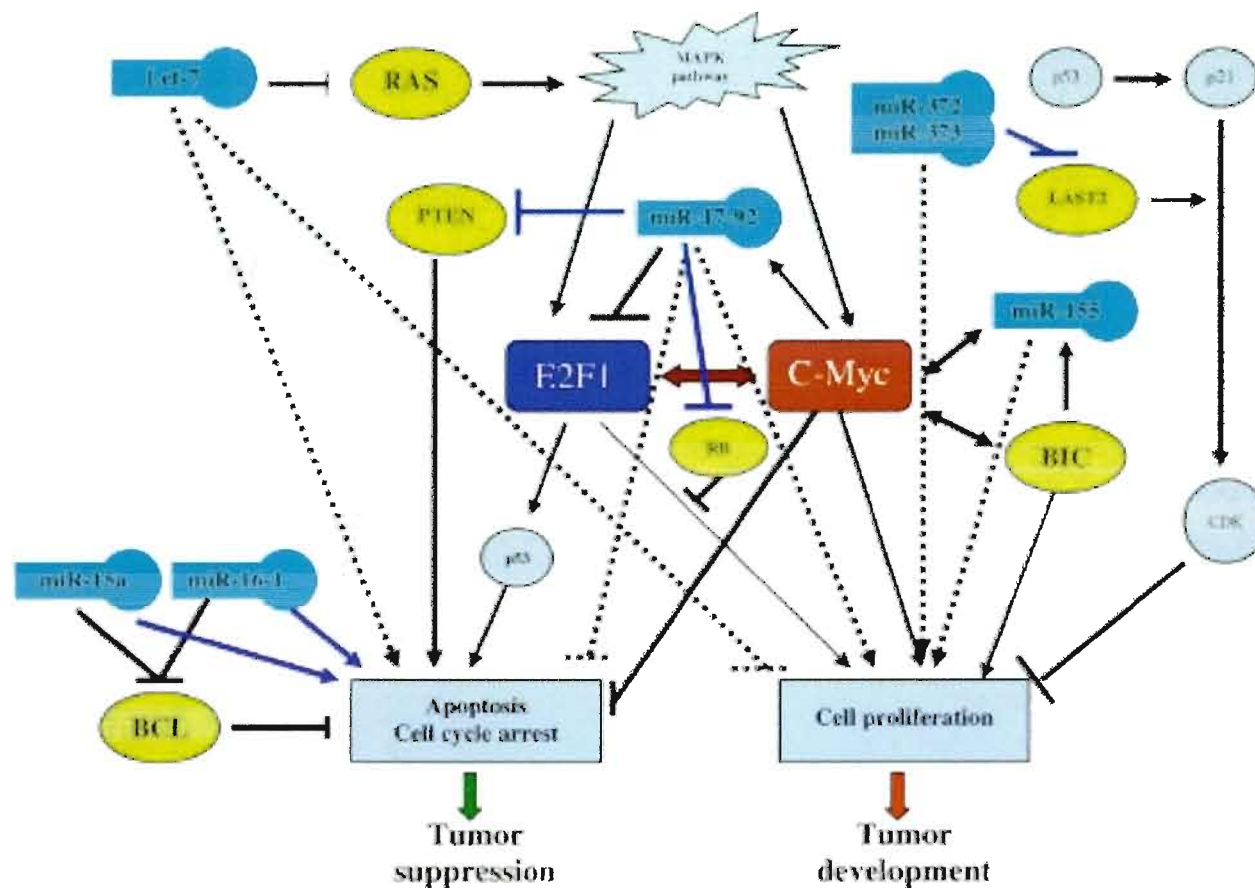
tissus cancéreux exhibent une diminution ou une augmentation d'expression de bon nombre de miARN. De façon surprenante, 50% des miARN connus se trouvent dans des régions génomiques associées au cancer telles que des zones de cassures chromosomiques ou des zones de délétion (Calin *et al.* 2004). L'hypothèse la plus évidente concernant le rôle des miARN dans le cancer est que ces derniers ciblent des oncogènes et des suppresseurs de tumeurs. Ainsi un miARN surexprimé dans un cancer exercerait son pouvoir oncogénique en ciblant des gènes suppresseurs de tumeurs, alors qu'un miARN sous-exprimé aurait plutôt un rôle de tumeur suppresseur par ciblage d'oncogènes (Fig. 1.9).

En combinant les analyses pour différents types de tumeurs solides (poumon, sein, colon, estomac, prostate et pancréas), Volinia *et al.* (2006) ont mis en évidence que 21 miARN étaient impliqués dans plusieurs types de cancers à la fois. Par exemple, miR-21 est impliqué dans les 6 types de cancers cités, alors que miR-17-5p et miR-191 le sont dans 5 de ces cancers. Ceci suggère donc que certains miARN sont impliqués dans des processus communs à tous les cancers (ou presque), et que d'autres sont plus spécifiques d'un type de cancer en particulier (Tableau 1.3).

### 1.2.3.2 MicroARN et leucémies

Comme dans le cas de cancers à tumeurs solides, les analyses de micropuces d'expression de miARN dans différents types de leucémies ont révélé de nombreux bouleversements (Calin *et al.* 2004b, Calin *et al.* 2005, Yu *et al.* 2006).

Le premier exemple de miARN impliqués dans le cancer concerne justement deux miARN impliqués dans la leucémie, miR-15a et miR-16-1, tous deux sous-exprimés chez la plupart des patients atteints de leucémies lymphocytiques chroniques (LLC, Calin *et al.* 2002). Ces deux miARN sont produits à partir de l'intron d'un ARNnc de fonction inconnue appelé Deleted in Lymphocytic Leukemia 2 (*DLEU2*), car ce gène est effectivement localisé dans une région de 30 Kb du chromosome 13 délétée dans plus de 65% des LLC. Une des cibles associées à ces deux miARN est *BCL2*, dont la répression permet l'apoptose (Cimmino *et al.* 2005). MiR-15a et miR-16-1 sont donc des exemples de miARN suppresseurs de tumeurs (Fig. 1.9). Par la suite, d'autres miARN ont été impliqués dans les leucémies, tels que le groupe *miR-17-92* (He 2005) qui est surexprimé dans des lymphomes



**Figure 1.9 Illustration des mécanismes moléculaires impliquant les miARN dans le cancer (extrait de Zhang *et al.* 2006).**

Dans cette figure let-7, miR-15a, miR-16-1 fonctionnent comme suppresseurs de tumeurs, alors que miR-17-92, miR-155, miR-372 et miR-373 sont considérés comme oncogènes. Les pointillés représentent des interactions indirectes. Les lignes noires représentent des interactions confirmées expérimentalement, alors que les lignes bleues représentent les interactions non confirmées. Abréviations : BCL2, B cell Lymphoma 2, antiapoptotic gene ; BIC : ARNnc hôte de miR-155; CDK : Cyclin Dependent Kinase; c-Myc : oncogène; E2F1 : a cell cycle transcription factor; LAST2 : Large Tumor Suppressor homolog 2 (sérine-thréonine kinase); MAPK :Mitogen-Activated Protein Kinase; p53 : tumeur suppresseur majeur; PTEN : Phosphatase and Tensin Homology, suppresseur de tumeur; RAS : oncogène.

Tableau 1.3 Exemple de miARN impliqués dans divers types de cancers (extrait de Garzon *et al.* 2006)

MicroRNA	Genomic location	Expression in patients	Experimental data	Function
MIR-15a	13q14	Down in CLL and in pituitary adenoma	Downmodulate BCL-2	TS
MIR-16-1				
Let-7a-2	11q24	Down in lung cancer	Downmodulates RAS and induces apoptosis in lung cancer cell lines	TS
MIR-155	21q21	Up in bad prognosis CLL and lung cancer Up in breast cancer, lymphoma Hodgkin and pediatric BL	Induces pre-B lymphoma and/or leukemia in mice	OG
MIR-17-92 cluster	13q14	Up in lymphomas and lung cancer	Cooperates with c-MYC, modulates E2F1	OG
MIR-21	17q23	Up in pancreas, glioblastoma and breast cancer	Anti-apoptotic in glioblastoma	OG
MIR-106a	xq26	Up in lung, gastric and prostate cancer	Downregulates RB-1	OG
MIR-372	19q13	Up in testicular germ-cell tumor cell lines	Neutralize p53 function	OG
MIR-373				
MIR-142	17q22	t(8;17) c-MYC is translocated downstream of the miR-142 hairpin, resulting in B-lymphoma	Enhances MYC expression	OG



de type B. Le rôle de miR-17-92 dans la leucémie semble complexe. D'une part ce groupe de miARN a été impliqué comme tumeur suppresseur de par son ciblage de E2F1, et sa capacité d'inhiber la prolifération cellulaire induite par c-myc (O'Donnell 2005). Mais d'autre part, *miR-17-92* a aussi été impliqué comme oncogène en inhibant l'apoptose induite par c-myc (He 2005). Or, rappelons ici que *c-myc* stimule la transcription de miR-17-92. Plus récemment, un rôle dans l'angiogenèse lui a été attribué : *miR-17-92* inhibe l'expression de deux protéines aux propriétés anti-angiogénique, Tsp1 et CTGF, et constitue ainsi le lien manquant entre c-myc et la sous-expression de ces deux protéines (Dews 2006). Quoiqu'il en soit, la dualité et la complexité de *miR-17-92* vis à vis du cancer n'a d'égale que celle de *c-myc*, auquel il semble intimement lié.

Récemment, miR-155 a été impliqué dans le développement de lymphomes de type B, et constitue un exemple intéressant de miARN associé à un ARNnc jusqu'ici mystérieux (Eis *et al.* 2005, Kluiver *et al.* 2005, Costinean *et al.* 2006). En effet, miR-155 est produit à partir d'un intron de l'ARNnc *BIC*, préalablement identifié comme oncogène associé à un site commun d'intégration du rétrovirus ALV (*Avian Leukosis Virus*) (Clurman *et al.* 1989, Tam *et al.* 1997). La surexpression de *BIC* était connue pour collaborer avec c-myc dans les lymphomes de type B ainsi que les érythroleucémies, sans que l'on sache de quelle manière l'ARNnc *BIC* agissait (Tam *et al.* 2002). La découverte de miR-155 lève donc le voile sur l'implication de *BIC* dans la leucémie.

Pour finir, citons miR-221 et miR-222 qui sont capables de cibler l'ARNm du récepteur *Kit*, et dont la sous-expression participe au développement d'érythroleucémies (Felli *et al.* 2006). Ou encore, la sous-expression de miR-29 et miR-181 dans les LLC de type B qui est associée à l'augmentation de leur cible, l'oncogène *TCL1* (Pekarsky *et al.* 2006).

### 1.2.3.3 Typage et pronostic tumoral par profilage d'expression des microARN.

De façon intéressante, les études par micropuce ont révélé des profils d'expression de miARN spécifiques d'un type de tumeur (pour une revue, voir Calin *et al.* 2006). Cela signifie qu'à terme, le diagnostic tumoral pourrait se faire par profilage de l'expression des miARN des patients. Ceci s'avère de première importance dans les tumeurs dont l'origine est incertaine, et de diagnostic difficile. C'est le cas des cancers métastatiques appelés CUP

(Cancer of Unknown Primary site) : les patients présentent des tumeurs métastatiques souvent très immatures, donc difficiles à classer, et dont on ne connaît pas le site d'initiation. Le typage par miARN de ce type de tumeurs s'est avéré plus fidèle et informatif que le typage par ARNm (Lu *et al.* 2005). De plus, une centaine miARN suffisent, au lieu de milliers d'ARNm.

Un autre exemple intéressant est celui des LLC. Ce type de leucémie couramment diagnostiquée est souvent associée à un pronostic favorable. En conséquence, si ces leucémies évoluent de manière agressive, les traitements ne sont administrés que tardivement. Le profilage des miARN de ces leucémies a permis la découverte d'une signature unique de 13 miARN capables de discriminer un bon pronostic d'un mauvais (Calin *et al.* 2005). Ceci permettrait donc l'administration d'un traitement préventif dans le cas d'une signature miARN associée à un pronostic défavorable.

Le profilage des miARN dans les cancers s'annonce donc comme un outil sans précédent pour la compréhension du cancer et l'ajustement thérapeutique.

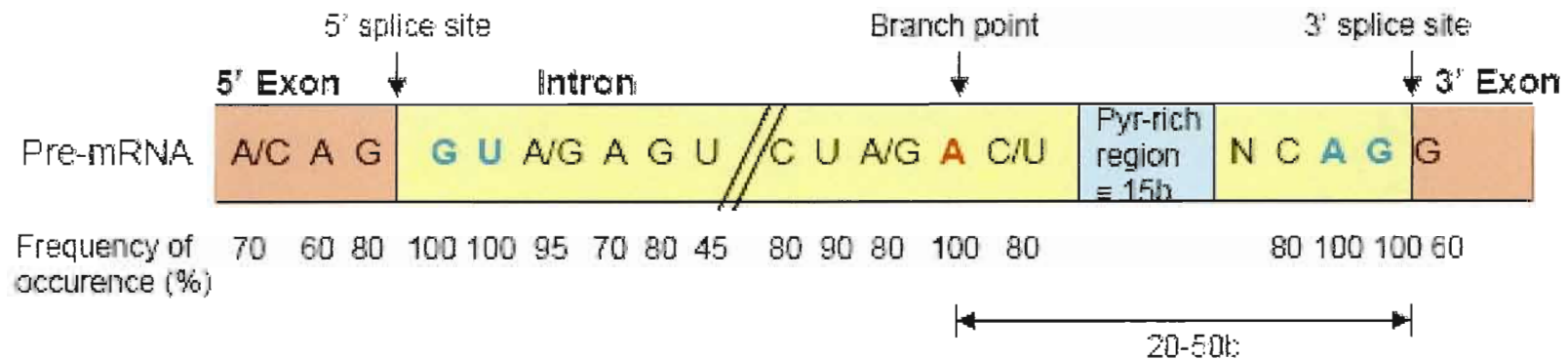


### 1.3 RÉGULATION POST-TRANSCRIPTIONNELLE DE L'EXPRESSION DES GÈNES PAR ÉPISSAGE ALTERNATIF

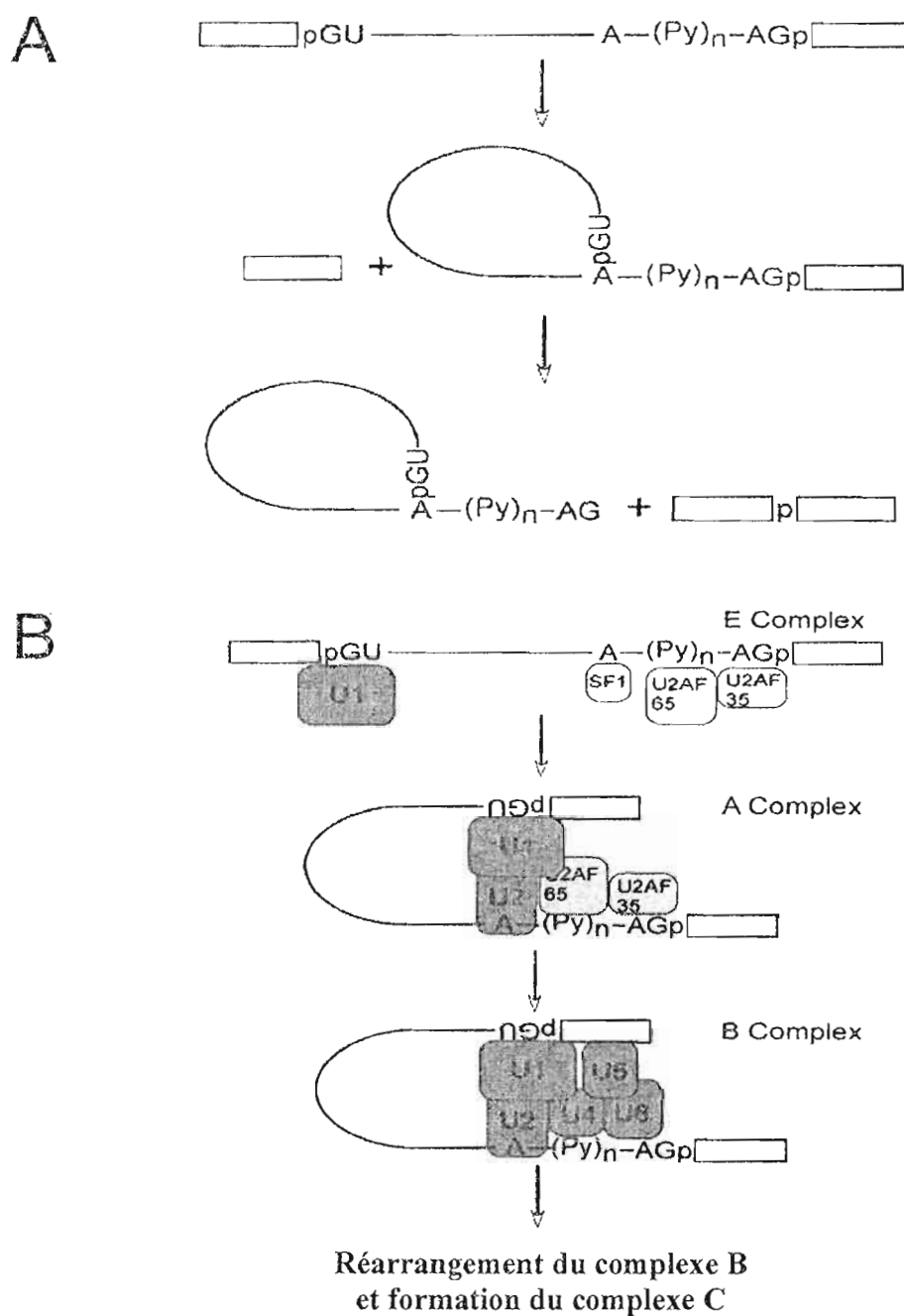
La plupart des ARNm subissent l'épissage de séquences introniques/exoniques de façon constitutive ou alternative. Ce processus nucléaire requiert une machinerie protéique appelée le spliceosome, mais également de nombreuses autres protéines non-spliceosomales (Kramer *et al.* 1996). L'épissage des ARNm est non seulement nécessaire à une bonne transmission du message tant structural que protéique, mais il doit aussi fournir les transcrits alternatifs appropriés en fonction des tissus et de leur développement. En effet, il est maintenant admis que l'épissage alternatif est un facteur clé de la diversité protéique. Le séquençage complet du génome humain a révélé l'existence d'à peu près 30 000 gènes, ce qui est loin d'être suffisant pour combler tous les mécanismes biologiques qui régissent des organismes aussi complexes que les mammifères. L'épissage alternatif apparaît donc comme une solution à ce paradigme : un gène peut ainsi être à l'origine de plusieurs protéines aux diverses fonctions. Ainsi, plus de 70% des gènes chez l'humain seraient soumis à des épissages alternatifs (Johnson *et al.* 2003). Un exemple incroyable de la diversité engendrée par l'épissage alternatif est celui du gène *Dscam* chez la Drosophile. Ce gène code pour un récepteur impliqué dans les connexions neuronales et peut être à l'origine de 38 016 transcrits différents, ce qui représente plus du double du nombre de gènes contenus dans le génome de la Drosophile (Black *et al.* 2000).

#### 1.3.1 MÉCANISME GÉNÉRAL DE L'ÉPISSAGE

L'excision d'un intron suivi de la liaison de deux exons d'un pre-ARNm est dirigée par les sites d'épissages qui sont des séquences consensus situées au niveau de la jonction intron/exon (Fig. 1.10) (Burge *et al.* 1999). Il s'agit d'une part d'une séquence GU (site donneur d'épissage, SDE) située en 5' de l'intron, elle-même comprise dans une séquence plus longue et moins conservée. En 3' de l'intron se trouve la séquence AG (site accepteur d'épissage, SAE). En amont du SAE, il existe deux autres éléments de séquence clés : le site de branchement (*branch point*) et une séquence riche en pyrimidines (*polypyrimidin track*).



**Figure 1.10 Séquences consensus d'épissage** (extrait de «Molecular biology of the cell», Albert 2002). Le site donneur d'épissage (SAE) en 5' de l'intron est en bleu (GU), ainsi que le site accepteur d'épissage (SAE) en 3' de l'intron (AG). Le résidu adénine essentiel dans le point de branchement est représenté en rouge. Pyr-rich region = région riche en pyrimidine.

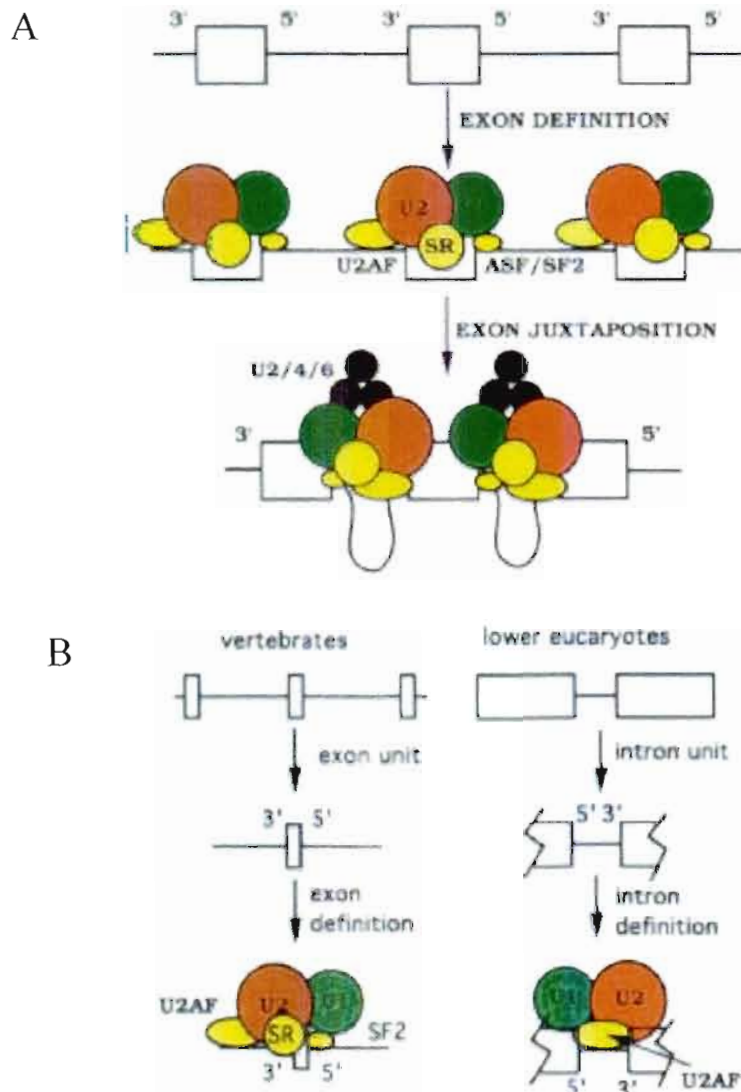


**Figure 1.11 Mécanisme général de l'épissage (extrait de Black *et al.* 2003).** (A) Les deux réactions de transestérification qui caractérisent l'épissage. (B) Assemblage du spliceosome. Voir texte pour les détails.

L'épissage peut se résumer à deux réactions de transestérification (Black 2003) (Fig. 1.11A). La première concerne l'attaque d'un phosphate du SDE par le groupement 2'OH du résidu adénosine du point de branchement. Ceci se solde par le détachement de l'exon en 5' de l'intron et la liaison du SDE au point de branchement, dans une conformation dite de «lariat». Dans la deuxième réaction, le 3'OH de l'exon précédemment libéré attaque le phosphate du SAE, pour créer une liaison avec l'exon en 3' de l'intron. Ce dernier est relargué, toujours sous la forme lariat, et les deux exons sont réunis.

L'épissage est réalisé par le spliceosome, composé de cinq ribonucléoprotéines dites «spliceosomales» (*small nuclear RNP*), et tout un ensemble de protéines accessoires qui encore aujourd'hui ne sont pas toutes connues. Brièvement, la snRNP U1 commence par se fixer au niveau du SDE en 5' de l'intron (Fig. 1.11B). En 3' de l'intron, plusieurs protéines interviennent : SF1, qui se lie au point de branchement, et la sous-unité de 65 kDa du facteur U2AF se lie au niveau de la séquence riche en pyrimidines. Certaines fois, la sous-unité de 35 kDa de U2AF se fixe au niveau du SAE. Ce premier ensemble de protéines constitue le complexe E (E pour Early). La snRNP U2 se lie au complexe E au niveau du site de branchement, pour former le complexe A, qui est à son tour rejoint par la tri-snRNP U4/U5/U6. Ce nouveau complexe est alors appelé complexe B, lequel va subir de nombreux réarrangements protéiques pour former le complexe C : entre autres, U1 et U4 disparaissent du complexe, alors que U6 prend la place de U1 au niveau de l'interaction avec le SDE. C'est le complexe C qui catalyse les deux réactions de transestérification décrites plus haut et qui caractérisent l'épissage.

D'une façon générale, les exons des ARNm chez les vertébrés sont minuscules par comparaison à la longueur des introns qui peut atteindre plusieurs milliers de paires de bases. On s'attendrait donc à ce que les séquences consensus qui gouvernent l'épissage soient spécifiques et très conservées, de façon à ce que la machinerie d'épissage les reconnaissent facilement. Or, ce n'est pas le cas : d'une part, les séquences consensus ne sont pas conservées, et d'autre part, elles sont assez répandues dans les introns. La question se pose donc de savoir comment la cellule fait la différence entre tous ces «pseudo-sites» d'épissage, qui se ressemblent plus ou moins. Une façon de résoudre ce problème pour les vertébrés serait de considérer la définition des exons plutôt que celle des introns (Berget 1995). Les snRNP U1 et U2 initiatrices de l'assemblage du spliceosome se fixeraient bien dans l'intron,



**Figure 1.12 La définition lors de l'épissage (Extrait de Berget 1995).** (A) Définition d'exons. Le spliceosome commence par s'assembler au niveau des sites d'épissage de part et d'autre de chaque exon, pour définir les exons à inclure dans le transcrit. Les protéines de type SR (telles que ASF/SF2) assurent la connexion entre les complexes. (B) Définition d'exons versus définition d'introns. Dans les cas où les introns sont de petites tailles (chez la levure par exemple), l'épissage se fait par définition des introns, par opposition à la définition d'exons (chez les vertébrés) lorsque les introns sont de grandes tailles.

mais en 3' de l'un et en 5' du suivant, de part et d'autre de l'exon (Fig. 1.12A). C'est ce que l'on appelle l'épissage par définition d'exon. De façon intéressante, c'est le contraire chez la levure et la *Drosophile* dont les gènes possèdent de petits introns d'une centaine de paires de bases, et des exons beaucoup plus grands (Zhang *et al.* 1994, Mount *et al.* 1992). Dans ce cas-ci, l'épissage est basé sur la définition des introns (Fig. 1.12B). Le type d'épissage, par définition d'exons ou d'introns, serait donc fonction de la taille de ces derniers. Ainsi chez les vertébrés, l'épissage pourrait aussi se faire par définition d'introns lorsqu'ils sont occasionnellement de petite taille.

### 1.3.2 ÉPISSAGE ALTERNATIF

L'épissage alternatif se traduit comme une variation dans le profil constitutif de définition d'exons (ou d'introns) d'un gène. Ces modifications sont gérées par une combinaison complexe de facteurs positifs et négatifs, dont les ratios sont importants pour la détermination des profils d'épissage. L'ensemble de ces facteurs peuvent être exprimés dans des conditions propres à chaque cellule, et constituerait un «code cellulaire» responsable de l'épissage alternatif spécifique d'un tissu ou d'un stade de développement, par exemple.

La régulation de l'épissage alternatif est assurée par des éléments de séquence *cis* et des facteurs *trans* (protéines ou ARN). Les séquences *cis* impliquées peuvent soit favoriser l'épissage (exon ou intron *splicing enhancers*, ESE ou ISE respectivement), soit l'inhiber (exon ou intron *splicing silencers*, ESS ou ISS) (Matlin *et al.* 2005). Les facteurs qui reconnaissent ces séquences ne sont pas forcément distincts de ceux impliqués d'ordinaire dans la définition d'exons/introns constitutive d'un gène. C'est plutôt la combinaison de différents facteurs qui orientera vers la favorisation ou l'inhibition de l'épissage. Les sites d'épissages peuvent aussi être qualifiés de forts ou de faibles, selon que leurs séquences soient 100% identiques ou non aux séquences consensus. Mais la seule «force» d'un site d'épissage n'est pas toujours suffisante pour imposer une définition d'exon. Un site d'épissage faible peut parfois être épissé plus efficacement, car il bénéficie d'une meilleure définition grâce au support d'ESE/ISE. La combinaison de ces facteurs gouverne l'épissage

alternatif d'un gène, qui se solde communément par une exclusion d'exon, une intrusion d'intron, ou le raccourcissement/allongement d'un exon (Wang *et al.* 2006).

### 1.3.2.1 Les séquences *enhancer* d'épissage (ESE et ISE)

Une activation d'épissage pour un exon se solde par l'inclusion de ce dernier dans le transcrit. L'exon est en quelque sorte préférentiellement «défini», que se soit grâce à des ESE ou des ISE (Fig. 1.13B).

Des études ont montré que les ESE sont de courtes séquences (5 à 7 nucléotides), hautement dégénérées et assez communes. Il est donc très difficile de les localiser avec précision. Ces séquences sont impliquées aussi bien dans la définition d'exons constitutifs d'un gène que lors de son épissage alternatif. Les ESE sont le plus souvent reconnues par les protéines SR (*Arginin and Serine rich proteins*). D'ailleurs, une protéine comme U2AF, facteur constitutif d'épissage, est une protéine SR (Black *et al.* 2003). Lorsque les protéines SR sont recrutées par les ESE, elles constitueraient notamment un lien entre U2AF au niveau de la séquence polypyrimidique en 3' de l'intron et le site d'épissage en 5', par de multiples interactions protéines-protéines (Shen *et al.* 2004).

Les ISE sont également assez variables en séquences, et la littérature est truffée d'exemples particuliers. Les ISE peuvent aussi être reconnues par des protéines SR mais, la plupart du temps, les protéines qui les reconnaissent sont moins bien caractérisées que dans le cas des ESEs (Black *et al.* 2003). Un exemple commun d'ISE est l'hexanucléotide UGCAUG qui, lorsqu'il est dupliqué, active fortement l'épissage. Il est le plus souvent dupliqué à proximité et en aval de l'exon qu'il active, mais peut aussi agir à distance (> 500 pb). L'épissage du gène c-src est par exemple activé par ce type d'ISE (Modafferi *et al.* 1997).

### 1.3.2.2 Les séquences *silencers* d'épissage (ESS et ISS)

Lors de l'inhibition de l'épissage d'un exon, c'est comme si ce dernier ne pouvait pas être reconnu par le spliceosome, et il est donc exclu du transcrit (Fig. 1.13A). Par un jeu de

blochage des sites d'épissage, des introns peuvent, au contraire, être inclus dans l'ARNm (Fig. 1.13C).

Contrairement aux ESE, les éléments de séquences constituant les ESS sont plus clairement identifiés. Les mieux caractérisés interagissent avec les protéines de la famille hnRNP (*heterogeneous Ribonuclear Protein*), qui lient spécifiquement l'ARN (Krecic *et al.* 1999). La plus étudiée de ces protéines est hnRNP A1, connue pour inhiber l'épissage via des ESS de nombreux transcrits, dont VIH (Virus de l'immunodéficience Acquis) (Caputi *et al.* 1999). Plusieurs mécanismes ont été proposés pour le mode d'action de hnRNP A1, dont le blocage de l'assemblage du spliceosome ou des interactions ESE/protéines SR (Black *et al.* 2003).

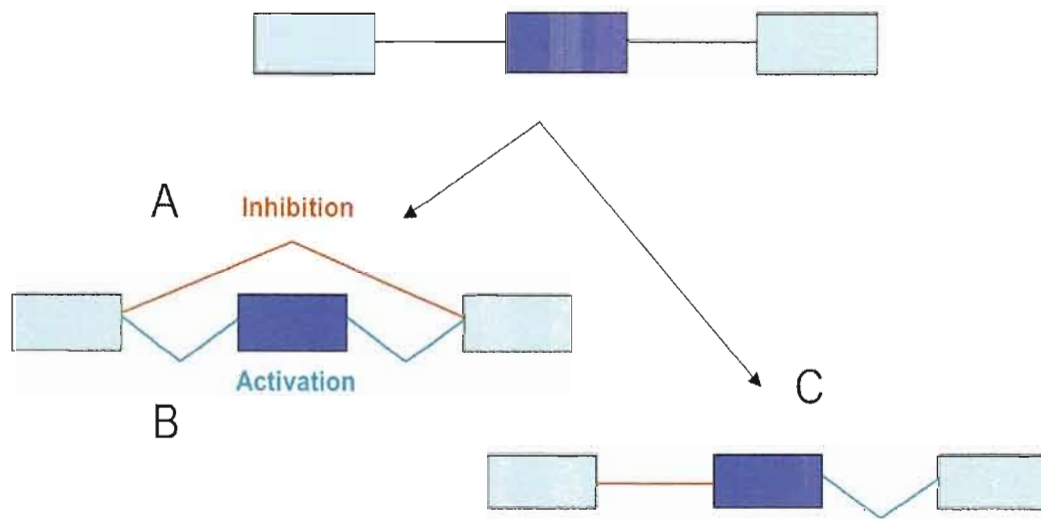
Les ISS peuvent être reconnues par des protéines SR ou même hnRNP A1. Une des protéines les plus connues pour inhiber l'épissage via les ISS est PTB (*Polypyrimidin Track Binding protein*), aussi appelée hnRNP I (Black *et al.* 2003). Ces sites de liaisons à l'ARN sont de type UUCU, mais de nombreux autres arrangements en pyrimidines (particulièrement riches en U) font aussi l'affaire (Singh *et al.* 1995, Perez *et al.* 1997). Ces arrangements nucléotidiques sont notamment retrouvés dans la séquence riche en pyrimidines en amont du site 3' d'épissage. Un des mécanismes d'action de PTB serait donc de se fixer au niveau de cette séquence et d'empêcher la fixation de U2AF. PTB est notamment impliquée dans la définition d'exon spécifique à un tissu donné en inhibant son inclusion dans les autres tissus (Black *et al.* 2003).

### 1.3.2.3 Combinaison de facteurs dans la régulation de l'épissage.

La définition de l'épissage alternatif est souvent le résultat d'une combinaison entre éléments régulateurs, positifs ou négatifs. L'activation ou l'inhibition dépendent la plupart du temps d'un équilibre entre les différents facteurs. Un simple changement dans cet équilibre peut profondément influencer sur l'épissage alternatif de certains transcrits, comme c'est par exemple le cas de la combinaison hnRNP A1 et SF2/ASF (Caceres *et al.* 1994).

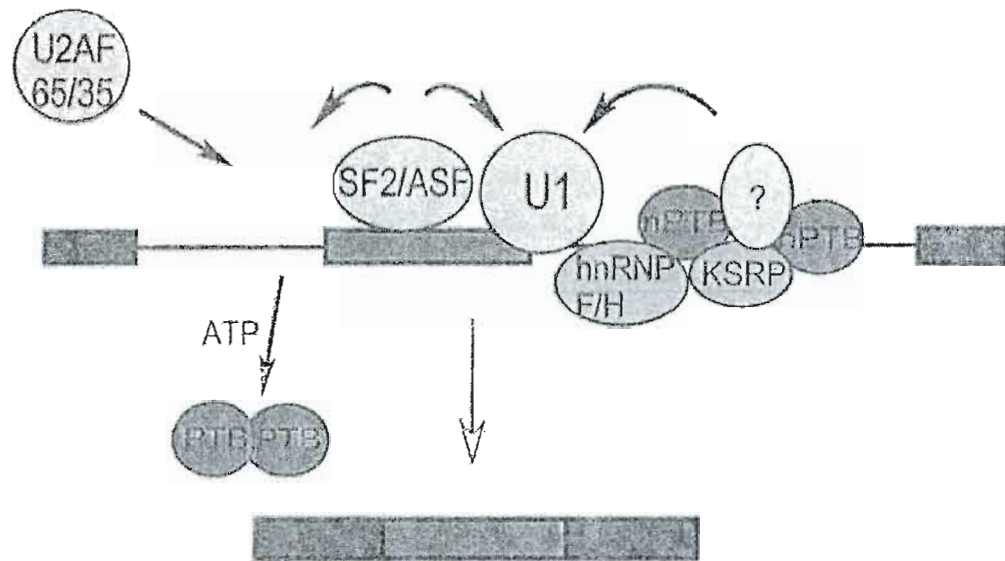
Pour illustrer la complexité relative à la régulation de l'épissage, j'ai choisi de détailler ici le cas du gène *c-src* (Fig. 1.14). Le gène *c-src* possède un exon spécifique aux neurones, appelé exon N1. Dans tous les autres tissus de l'organisme, N1 est exclu du transcrit (Black *et al.* 1992, Modafferi *et al.* 1997, Black *et al.* 2003). L'épissage de N1 est



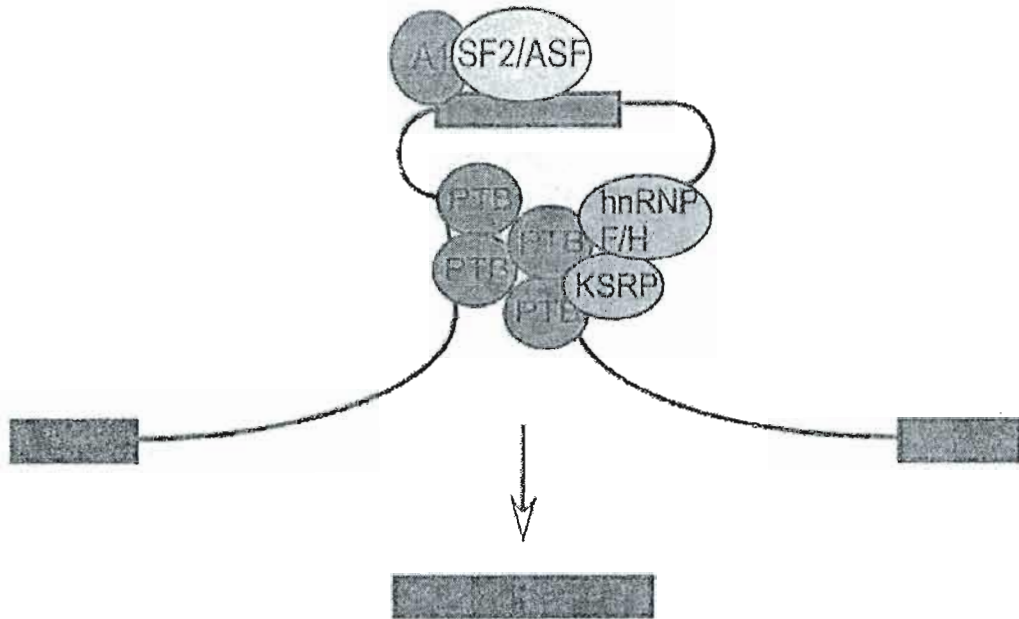


**Figure 1.13 Épissage alternatif.** (A) Les sites d'épissage de part et d'autre de l'exon épissé alternativement (bleu foncé) sont inhibés. L'exon est donc exclu (appelé *exon skipping*). (B) Les sites d'épissage de part et d'autre de l'exon sont activés et l'exon est donc inclus dans le transcrit. (C) Rétention d'intron. Les sites d'épissage de part et d'autre de l'intron (en rouge) sont inhibés, alors que le site d'épissage en 3' de l'exon alternatif est activé.

gouverné par plusieurs éléments. Tout d'abord, dans l'intron en aval de ce dernier exon se trouve un élément régulateur, communément appelé DCS (*Downstream Control Sequence*). Il contient notamment trois éléments de séquences distincts : GGGGG et UGCAUG qui sont des ISEs, et l'ISS CUCUCU. Le DCS est un point d'ancrage pour un large complexe de protéines liant l'ARN, qui sont en partie connues (Markovstov *et al.* 2000). L'ISE GGGGG fixe hnRNP F et H, alors que l'ISE UGCAUG fixe la protéine KSRP. Quant à l'ISS CUCUCU, il lie la protéine PTB, ou son homologue neuronale nPTB. Dans les cellules non neuronales, PTB est fixée à des séquences introniques en amont de N1, ainsi qu'en aval au niveau du DCS. Selon l'hypothèse, cette répartition de PTB serait à l'origine d'une boucle dans l'ARN de c-src au niveau de N1 qui séquestrerait ce dernier et inhiberait sa sélection lors de l'épissage (Fig. 1.14). Les protéines KSRP et hnRNP F/H sont également présentes au niveau du DCS avec PTB. Dans les neurones, nPTB prend la place de PTB au niveau du DCS, toujours en association avec KSRP et hnRNP F/H, et lève l'inhibition d'épissage sur l'exon N1. Par un mécanisme pour l'instant inconnu, PTB est dissociée de ces éléments de liaison par une activité ATP dépendante. Pour rajouter encore à la complexité de cette régulation, des sites de liaison pour les protéines hnRNP A1 et ASF/SF2 sont présents en 5' dans l'exon N1. Il a été effectivement montré que ces deux protéines agissent de concert dans la régulation de l'épissage de N1, avec ASF/SF2 exerçant un rôle positif, et hnRNP A1 un rôle négatif (Rooke *et al.* 2003). L'épissage alternatif de l'exon N1 de c-src est donc à la fois, et indépendamment, contrôlé par une combinaison d'ISE/ISS et une combinaison d'ESE/ESS.



## Neurons



## Non-neuronal Cells

Figure 1.14 Épissage alternatif de l'exon N1 du gène c-src (extrait de Black 2003). L'exon N1 est représenté en gris pâle. Voir texte pour explications.

## CHAPITRE II

### DÉCOUVERTE ET CARACTÉRISATION DU SITE COMMUN D'INTÉGRATION *KIS2*

#### AVANT-PROPOS

Ce chapitre correspond à la première partie de ma thèse, qui se solde par la publication dans "Journal of Virology" d'un premier article intitulé "Radiation leukemia virus common integration at the *Kis2* locus : simultaneous overexpression of a novel non-coding RNA and of the proximal Phf6 gene". (Landais 2005)

Cet article rapporte la découverte d'un nouvel SCI localisé sur le chromosome X, que nous avons appelé *Kis2*. Grâce à la publication du séquençage du génome de la souris, il a été possible de répertorier tous les gènes de cette région. Nous avons trouvé que deux gènes étaient surexprimés dans les tumeurs possédant une intégration au niveau de *Kis2*. L'un des gènes, que nous avons baptisé *Kis2*, est un nouvel ARN non-codant au profil d'épissage complexe. Le deuxième gène est situé 300 kb en amont des intégrations et correspond à Phf6. Jusqu'à présent, aucun de ces deux gènes n'a été impliqué dans un quelconque processus tumoral.

Lors de ce travail, Renaud Quantin a mis au point et réalisé la technique de PCR inverse, puis le clonage des intégrations rétrovirales. J'ai par la suite mis en évidence que le site d'intégration *Kis2* était un SCI et réalisé sa caractérisation. J'ai moi-même réalisé intellectuellement et techniquement l'intégralité des travaux, et rédigé le manuscrit.



RADLV COMMON INTEGRATION AT THE *KIS2* LOCUS :  
SIMULTANEOUS OVEREXPRESSION OF A NOVEL NON-CODING RNA  
AND OF THE PROXIMAL *PHF6* GENE.

SÉVERINE LANDAIS, RENAUD QUANTIN AND ERIC  
RASSART\*

Laboratoire de biologie moléculaire, Département des sciences biologiques,  
Université du Québec à Montréal, Québec, Canada.

Running title: Common integration site in RadLV-induced tumors

**KEY WORDS:** Radiation leukemia retrovirus, Common integration site,  
Oncogene activation, *Phf6*, non-coding RNA

\* Correspondance to: Eric Rassart

Département des Sciences Biologiques

Université du Québec à Montréal

Case Postale 8888 Succ. Centre-ville

Montréal, Canada . H3C-3P8

Tel. (514)987-3000 ext : 3953. Fax. (514)987-4647.

Email : Rassart.Eric@UQAM.ca

## 2.1 ABSTRACT

Retroviral tagging has been extensively and successfully used to identify genes implicated in cancer pathways. In order to find oncogenes implicated in T-cell leukemia, we used the highly leukemogenic VL3 Radiation Leukemia Retrovirus (RadLV/VL3). We applied the inverted PCR technique to isolate and analyze sequences flanking proviral integrations in RadLV/VL3-induced T-lymphomas. We found retroviral integrations in *c-myc* and *Pim1* as already reported but we also identified for the first time *Notch1* as a RadLV common integration site (CIS). More interestingly, we found a new RadLV CIS that is situated on the mouse chromosome X (XA4 region, 45,091,000 bp). This site has also been reported as an SL3-3 and Mo-MuLV integration site, which strengthens its implication in MuLV induced T-lymphomas. This locus, named *Kis2* (Kaplan Integration Site 2), was found rearranged in 11 % of the tumors analysed.

In this article, we report not only the alteration of the *Kis2* gene located nearby in response to RadLV integration but also the induction of the *Phf6* gene expression situated about 250 kb from the integration site. The *Kis2* gene encodes 5 different alternatively spliced non-coding RNAs and the *Phf6* gene codes for a 365 amino acids protein which contains two Plant Homology Domain (PHD) fingers, recently implicated in the Boffi esonForssman-Lehmann Syndrome in human.

With the recent release of the mouse genome sequence, high-throughput retroviral tagging emerges and provides a powerful tool for the oncogene quest. It also allows for the analysis of large DNA regions surrounding the integration locus.

## 2.2 INTRODUCTION

Cancer is a complex disease that involves a cascade of events and the collaboration of several genes, including oncogenes and tumor suppressors. Several lines of evidence indicate that cell transformation is under the control of a few major pathways (21, 22, 68). Several key oncogenes have been identified for each of these, such as c-myc and ras. Such key oncogenes are necessary although not enough to induce tumors, and other genes are required. Secondary events such as the activation of tumor progression genes or inactivation of tumor suppressor genes are necessary to maintain the tumoral state and to allow the appearance of the tumor. Many of those collaborating genes remain to be discovered.

Retrovirus-induced tumors are interesting models to study those different pathways. Indeed, retrovirus pathogenesis mimics classical events in the development of a tumor cell but with a shorter latency. The leukemogenic potential of nontransforming retroviruses is a direct consequence of the random integration of their proviral DNA genome in the host cell DNA during the replicative cycle. Retroviral integration sites, when common to several tumors, are considered significantly associated with tumor induction and/or development. As those loci generally contain oncogenes, retroviruses have been widely used over the years as molecular tools to identify cancer-associated genes. Recently, mouse genome sequencing demonstrated powerful use of retroviruses with the emergence of high-throughput retroviral tagging (30, 32, 45, 50, 68). When performed on a transgenic mouse model overexpressing a known oncogene or a mutated tumor suppressor gene, collaborating oncogenes from a particular pathway could be identified (27, 45, 50). All these data are now listed in the Mouse Retroviral Tagged Cancer Gene Database (<http://genome2.ncifcrf.gov/RTCGD>).

The radiation leukemia virus (RadLV) is a radiation-induced retrovirus that became leukemogenic after serial passages in mice (61). Originally thymotropic, weakly leukemogenic, nondefective, and ecotropic when isolated from X-ray-induced T-cell lymphomas in C57BL/6 mice (40, 42, 60), serial passages in newborn C57BL/Ka mice gave rise to a nonfibrotropic virus preparation with a very high leukemogenic potential (31, 40,



60). Several lymphoid cell lines were established from tumors induced by the passaged RadLV, including the BL/VL3 cell line (41). Several molecular clones were obtained from the BL/VL3 cell line as proviruses and several variants were characterized (59, 61). More particularly, the highly leukemogenic molecular clone V-13 provides an interesting T-cell lymphoma model, mostly because the disease latency is less than 3 months (60, 61).

Until now RadLV/VL3 has been rarely used in retroviral tagging. Previous studies report common integration near the *cyclin D2* gene (*Vin1* locus), *Pim1*, and *c-myc* in 5%, 30%, and 15%, respectively, of tumors (23, 70). *Pim1* and *c-myc* are both known to be involved in T-cell lymphomas (9, 64, 65, 71). Common integrations in the *Kis1* locus on mouse chromosome 2 have also been reported, but, so far, no gene has been associated with this locus (35).

In the present study, we used the CpG island technique of Li et al. (39) in an attempt to identify new proto-oncogenes in RadLV-induced T-cell lymphomas. We provide evidence of RadLV integration in the *Notch1* gene. We report the identification of a new common integration site named *Kis2* (Kaplan Integration Site 2) which is located on the mouse X chromosome. Retroviral integration was observed in 11% of the tumors and it modifies the expression of two genes in the region.

## 2.3 MATERIALS AND METHODS

### 2.3.1 Virus, mice and tumors

Several RadLV variants were used to generate leukemias. These are molecular clones from the BL/VL3 tumor cell line (61). The SIM/R and C57BL/6 mouse strains were obtained from Dr. Arthur Axelrad (Toronto, Canada) and from Charles River Canada (St-Constant, Quebec, Canada), respectively. Newborn mice were inoculated intraperitoneally with 0.15 ml of filtered virus suspension obtained from independent clones ( $10^5$ - $10^6$  virions/ml). Most mice developed tumors within 3-5 months postinoculation, and they were sacrificed when obvious lymphoid organ enlargement was detected through palpation. Mice from both strains had a typical enlarged thymus, often accompanied by enlarged lymph nodes and, more rarely, enlarged spleen (61). Tumors rearranged for *Kis2* are : T1: 255.3, T2: 132.9, T3: 139.1, T4: 255.6, T5: K31.5.

### 2.3.2 Genomic-DNA and RNA analysis

Genomic DNA was extracted from frozen normal and leukemic tissues by standard procedures as described previously (63). For southern analysis, 15  $\mu$ g of genomic DNA were digested overnight with suitable restriction enzyme, separated on a 0.8% agarose gel and transferred to a nylon membrane (Osmonics, Fisher, Ottawa, Ont., Canada).

Total RNA was extracted from frozen normal and leukemic tissues with the TRIzol reagent (Invitrogen, Burlington, Ont., Canada) according to the manufacturer's protocol. For Northern analysis, 15  $\mu$ g of total RNA was separated on a 1% formaldehyde agarose gel as described previously (63), and transferred to a nylon membrane (Osmonics).

All the membranes were prehybridized for 1h at 42°C in 50% formamide-5X SSPE (3M NaCl, 20mM EDTA, 20mM sodium phosphate pH 6.8)-5X Denhardt's-0.5%SDS,-0.1mg/ml denatured salmon sperm DNA. Hybridization was performed overnight at 42°C in

the same buffer containing 10% dextran sulfate. The probes were labelled by the random primer extension method using oligohexamers (GE Healthcare, Baie d'Urfé, Qc, Canada), and approximately  $2 \times 10^6$  cpm/ml were used. The membranes were washed for 10 min at 42°C in 2X SSC (1X SSC is 0.15 M NaCl-0.015 M sodium citrate), 30 min at 42°C in 2X SSC-0.1% SDS, and finally for 15 min at 65°C in 0.1X SSC-0.1% SDS. Membranes were revealed by exposure and scanning with a phosphorimager (BioRad, Mississauga, Ont., Canada).

### 2.3.3 RadLV integration site amplification by I-PCR

The viral insertion site amplification (inverse-PCR, I-PCR) was performed essentially according to the method of Li *et al* (39). Briefly, 2 ug of tumor DNA were digested with SacII and allowed to circularize with T4 DNA ligase (GE Healthcare). Due to the presence of a SacII site in the RadLV/VL3 gag region, DNA molecules that contained a provirus portion and flanking cellular sequences were specifically PCR amplified with a pair of oligonucleotides complementary to respectively the U3LTR and the gag region upstream of the SacII site.

Confirmation of retroviral integration at the *Kis2* locus was done by PCR with the Taq polymerase kit (Qiagen, Mississauga, Ont., Canada) according to the manufacturer's protocol. Briefly, 200 ng of genomic DNA from tumors were used as template in a 50 µl reaction, with 10 pmole of either U3 sense RadLV primer (U3624, 5'-TGGGCCCCGGCTCCGTTAGACATA-3') or U3 reverse RadLV primer (U3nsens, 5'-GGCTGGGACTTTCCAGAACTGTTG-3'), and 10 pmole of either *Kis2* genomic primer sense (255.4s, 5'-CATGCGCACTACCTCACAGCTAGAT-3') or *Kis2* genomic primer reverse (255.4ns, 5'-ATCTAGCTGTGAGGTAGTGCGCATG-3'). PCR program was one cycle of 94°C 2 min, and 30 cycles of 94°C 20 sec, 57°C 30 sec, 72 °C 5 min, and one cycle of 72°C 10 min.

### 2.3.4 5' and 3' RACE, CR-RT-PCR, and RT-PCR

Both 5' and 3' Race on *Kis2* RNAs were performed using the GeneRacer Kit (Invitrogen), according to the manufacturer's protocol. To isolate 3' ends, the RT reaction was primed with the oligo dT anchor of the Kit, and PCR performed with combination of the 3' anchor primer and a *Kis2* primer (oligo 2, 5'-GTGACATCTCTAGGCCATCTGTGCGGTATGTG-3'). To isolate 5' ends, the RT reaction was primed with a primer internal to *Kis2* (oligo D, 5'-CTATTCCCGCATCTTCAGCAGCGGCTCCAG-3'), and PCR performed with the primer specific to the 5' anchor and a nested *Kis2* specific primer (oligo IS, 5'-CAAAACGCGCCAAGCTTAAGCCTCTCCCTCCATTCTC-3'). PCR conditions for both 5' and 3' RACE were : 94°C 5 min, and 30 cycles of 94°C 30 sec, 63°C 30 sec, 72°C 1 min 30 sec, and 72°C 10 min.

Circularized RNA-RT-PCR (CR-RT-PCR) was performed as described in Kuhn (34). Briefly, 10 µg of total RNA from *Kis2* tumors were dephosphorylated, the cap was removed, and the resulting RNA was circularized by ligation. RT (Omniscript, Qiagen) was primed with two different *Kis2* primers, either oligo D or oligo IS described above followed by PCR with nested primers. PCR products were loaded on 1% agarose gel and the bands were eluted, cloned and sequenced.

RT-PCR for detection of *Phf6* in tumors was performed on 300 ng of total RNA as follows: RT was primed with an oligo dT and PCR was performed with primers g49rik1 (5'-CTGTCCCACGAACCTAGTTTCCATTAAG-3') and g49rik2 (5'-CAGGAAATCAAA CAGGTACCCAGGAAAATG-3'). PCR reaction was performed with 1/2 of the RT product reaction: 94°C 5 min, and 25 cycles of 94°C 30 sec, 62°C 30 sec, 72°C 1 min 30 sec, and 72°C 10 min.

Nested-RT-PCR to detect *Kis2* expression in normal tissues was performed as follows: total RNA from normal tissues extracted by TRIzol reagent were subsequently cleaned by phenol-chloroform extraction, to eliminate phenol contamination and ensure RT reproducibility. RT was performed with 300 ng of total RNA, and primed with the *Kis2* primer (oligo B 5'- GACATCGCCAAGGCATACTCCAGGAGTGTAACC-3'). The first PCR was done with oligo B and oligo IS4bis (5'- ATCGGGCTTTGGGACGCAGCTCGAG-

3'), and the following program was used: 94°C 5 min, and 15 cycles of 94°C 30 sec, 62°C 30 sec, 72°C 2 min, and 72°C 10 min. The PCR products were then cleaned on Qiaquick PCR purification column (Qiagen), and 1/30 of the eluate was used in the nested PCR performed with oligo IS4bis and a *Kis2* oligo nested to oligo B ( B2 5'- TCTTCCACGA ACTCTTCTTTGCCAGATCTGCAG-3'), following the PCR program 94°C 5 min, and 25 cycles of 94°C 30 sec, 64°C 30 sec, 72°C 2 min, and 72°C 10 min. The PCR products were analyzed on a 2% agarose gel.

Nested-RT-PCR to detect *Phf6* transcripts in normal tissues was performed as for *Kis2*. RT was primed with an oligo dT. Primers used in the first PCR were oligo 1-3R1 (5'-TGGAATTCTCGAAACTACCCCTGGCGACAG-3'), and oligo 5-4SmaI (5'-TACTGACCCGGGTGTTTCCATTAAGCTGCTG-3'). The two oligos contained an EcoRI site and a SmaI site, respectively. The PCR program was: 94°C 5 min, and 15 cycles of 94°C 30 sec, 56°C 30 sec, 72°C 1 min 30 sec, and 72°C 10 min. For the nested PCR, primers were oligo 1-3R1 and Phf6TCSma (5'-TACCCCGGGTCCACATTTAGGCCTGGTAT-3') but the PCR program was for 25 cycles. The PCR products were migrated on a 2% agarose gel.

The  $\beta$ actin gene was used as control in all the RT-PCR performed. The reverse primer was either oligo dT or an internal primer ( $\beta$ act2, 5'- GACGGGGTCACC CAACTGTGCCCATCTA-3'). PCR reaction was performed with 1/10 of the RT product reaction in presence of primers  $\beta$ act2 and  $\beta$ act1 (5'-CTAGAAGCACTTGCG GTGCACGATGGAGGG-3'). PCR program was: 94°C 5 min, and 25 cycles of 94°C 30 sec, 62°C 30 sec, 72°C 1 min, and 72°C 10 min.

### 2.3.5 cDNA cloning of *Phf6*

A *Kis2*-rearranged tumor cDNA library was prepared with the Superscript Plasmid System with Gateway Technology for cDNA synthesis and cloning kit (Invitrogen). The library was made using 4  $\mu$ g of polyA<sup>+</sup> RNA, and cDNA ligation product was precipitated prior to electroporation in electrocompetent bacteria DHB10 (Invitrogen). One  $\mu$ l of the ligation was used to transform 25  $\mu$ l of bacteria. A total of  $5 \times 10^5$  clones were plated on 22,5 x 22,5 cm L-Broth agar plates ( $8 \times 10^4$  clones per plate) containing 100  $\mu$ g/ml ampicillin,

grown overnight at 30°C, transferred on nylon membranes in duplicate and screened with a *Phf6* specific probe (RT-PCR product eluted from agarose gel). Positives clones were identified on the original plates, isolated and sequenced.

### 2.3.6 PHF6 localization in NIH3T3 cells

NIH3T3 cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% calf serum (Invitrogen). The DNA sequence corresponding to the protein encoded by transcripts t2 and t7 from *Phf6* was PCR amplified using primers 1-3R1 (5'-TGGAATTCTCGAAACTACCCCTGGCGACAG-3') and 5-4Sma1 (5'-TACTGACCCGGGTGTTTCCATTAAGCTGCTG-3') for Phf6-t7, and primers 5-4R1 (5'-TGGAATTCTCCTTAAAGTGGCAGTCTCTAAAC 3') and 5-4Sma1 for Phf6-t2. PCR products were eluted from agarose gel, and cloned in pEGFP-N1 and pEGFP-C2 (BD pharmingen, Mississauga, Ont., Canada). NIH3T3 cells were seeded at  $2 \times 10^5$  cells per plates and transfected with the different constructs mixed with polyfect reagent (Qiagen) according to the manufacturer's protocol. Fluorescent cells were analysed 24h post transfection by confocal microscopy.

### 2.3.7 Sequence analysis

DNA sequences were performed at Canadian Molecular Research Services Inc. ([www.cmrsinc.com](http://www.cmrsinc.com); Montreal, Qc, Canada). Sequence analysis was done with the BLAST program of both databases Ensembl and NCBI. Sequence alignment were performed by CLUSTAL.

## 2.4 RESULTS

### 2.4.1 Integration pattern in RadLV-induced tumors.

In an attempt to find new RadLV common integration sites, we used the CpG island technique described in Li et al. (39) to rapidly clone and sequence cellular DNA flanking the proviruses. We examined 72 tumors generated by inoculating newborn C57/Bl and SIM.R mice intraperitoneally with the RadLV/VL3 molecular clone. We found RadLV integrations in the *Pim1* and *c-myc* genes in 30% and 15% of the tumors analyzed, respectively (Fig. 2.1). Similar results have already been reported by Tremblay et al. (70). In our tumors, integrations in *Pim1* were all clustered in the 3' region of the gene, while in the case of *c-myc*, integrations occurred upstream of exon 1, in the promoter region of the gene. This integration pattern is similar to that observed in T-lymphoma-inducing retroviruses (9, 10, 65; reviewed in reference 12).

We also found RadLV integrations in the *Notch1* gene in 16% of the tumors analyzed. This gene encodes a transmembrane receptor that is important in hematopoiesis for stem cell commitment to develop into functional T cells and particularly for the assembly of pre-T-cell receptor complexes in immature thymocytes (20, 47). Activating mutations in the heterodimerization domain and/or in the terminal PEST domain of NOTCH1 seem to be implicated in 50% of human acute T-lymphoblastic leukemias (74). However, retroviral targeting of *Notch1* has not often been reported. Common retroviral integrations in *Notch1* have been reported in two cases, one being in T-cell lymphomas induced in BALB/c mice by a Moloney murine leukemia virus mutant carrying specific deletions in the U3 long terminal repeat region (76), and the other in tumors from transgenic mice infected with Moloney murine leukemia virus (15, 24). Other single hits in *Notch1* have since then been reported by high-throughput analysis (for a summary, see the Mouse Retroviral Tagged Cancer Gene Database).

For RadLV/VL3, integrations were clustered in the last exon of the gene. This corresponds to the type II integrations described by Hoemann et al. (24) although, contrary to their study, we found proviruses in both orientations. RadLV/VL3 integrations occur, in terms of protein, before the OPA motif (tumors 153.1 and 271.1), within the OPA motif (tumor 255.6), after the PEST motif (tumor 255.7), or 400 nucleotides downstream of the stop codon in the 3' untranslated region (tumor 137.2). *Notch1* activation in leukemia has been reported to be often associated with *c-myc* (15, 24) but none of our *Notch1*-rearranged tumors was also rearranged for *c-myc* (results not shown).

#### 2.4.2 *Kis2*, a new common integration site.

One of the RadLV/VL3 integration site corresponded in the database to a single SL3-3 integration site (67), suggesting that this region could be a common integration site. Southern blot and PCR analyses revealed retroviral integration in this specific locus in 11% of the tumors analyzed (8 tumors out of 72). This new common integration site is situated on the X chromosome, more precisely in a region corresponding to bp 45091000 (XA4 region) according to the mouse genome sequence database ([www.ensembl.org](http://www.ensembl.org)). We named this site *Kis2* (Kaplan Integration Site 2). Five of the tumors are almost clonal since the rearranged allele appears slightly weaker than the normal one (Fig. 2.2A and not shown). Note that the normal allele is absent in tumor 2 (T2) because this mouse was a male. DNA rearrangement in the three other tumors (T4, T7, and T8) was not detectable in Southern blots, but retroviral integration was easily detectable by PCR.

Southern blot analysis of DNA from RadLV/VL3-induced tumors with a U3 long terminal repeat probe or a T-cell receptor (TcR) probe has already been performed and these studies show that RadLV tumors are clonal cell populations (2, 35, 70). However, these studies did not allow to determine if the total cell population was homogenous. Our PCR-based strategy certainly allows to clone not only retroviral integrations that are present in the total tumor cell population but also in more minor cell population.



Figure 2.2B shows the gene and expressed sequence tag (EST) map of the *Kis2* region. We limited the region on one side by the glypican 3 gene (*Gpc3*), and by the *Phf6* gene on the other side. *Gpc3* encodes a heparan sulfate proteoglycan which is involved in developmental morphogenesis (13), while *Phf6* encodes a protein implicated in a mental disorder called Borjeman-Forsmann-Lehman syndrome (BFLS) (43). A third gene, *Rs17*, is present and encodes a ribosomal protein. Three ESTs (AI464896 and unigene clusters Mm. 277876 and Mm. 23411) have also been reported in this region (Fig. 2.2B). All RadLV/VL3 integrations were found clustered in a 4-kbp region between the *Rs17* gene and the Mm.277876 EST cluster. SL3-3 and Moloney murine leukemia virus integration sites are also present in this region.

Most of the time, retroviruses were integrated in an antisense orientation with respect to the transcription direction of the Mm.277876 EST cluster (Fig. 2.2B, tumors T1, T2, T3, T4, T7, and T8), as well as SL3-3. However, in the case of tumors T5 and T6, the retrovirus is integrated in the sense of transcription and closer to Mm.277876 (0.5 kbp). Interestingly, this region was also reported as a common integration site (called *XPCL1*) for Moloney murine leukemia virus in p27<sup>-/-</sup> transgenic mice (27). Moloney integrations are localized, although not oriented, to the same area as RadLV integrations (Fig. 2.2B), and overexpression of EST AI464896 was reported in the Moloney-rearranged tumors. This EST is located 4 kbp downstream of the RadLV and Moloney integrations, and 0.5 kbp downstream of the Mm.277876 cluster (Fig. 2.2B). However, EST AI464896 expression was not altered in our RadLV-induced tumors (not shown) suggesting that another EST or gene was targeted in our tumors.

Thus, we decided to analyze the expression of all the genes and ESTs present in the entire 450-kbp region. We did not find any change of expression for the *GPC3* and *Rs17* genes or for Mm.23411 EST. Interestingly, we found that expression of EST from the Mm.277876 cluster was drastically increased in the *Kis2* rearranged tumors. Therefore, we believe that the RadLV/VL3 retrovirus, although integrated in the same region as the Moloney murine leukemia virus, is targeting another gene. Also, the expression of the *Phf6* gene was affected (see below).

### 2.4.3 A novel gene is overexpressed in *Kis2*-rearranged tumors.

Each gene and EST present in the region of 450 kbp was tested by Northern or RT-PCR analysis to detect altered expression in the *Kis2*-rearranged tumors. Mm.277876 expression was drastically altered in all *Kis2*-rearranged tumors. We named the gene corresponding to the Mm.277876 EST cluster *Kis2*.

Two clearly different patterns of expression were obtained depending upon the orientation of the provirus. First, five transcripts of different sizes were detected in tumors T1, T2, and T3 (Fig. 2.3A) and second, the third RNA species was largely overexpressed in tumors T5 and T6, where the retrovirus is integrated in the sense orientation. Overexpression of *Kis2* transcripts was barely detectable in tumors T4, T7, and T8 due to the nonclonality of these tumors. In the nonrearranged tumors, the level of expression of *Kis2* transcripts remained very low. Interestingly, the BL/VL3 cell line which is a lymphoid cell line established from a tumor induced by the passaged RadLV, showed low but significant quantities of the three longest transcripts. Nevertheless, no integration could be found in the immediate 6-kbp region of the *Kis2* locus in this cell line.

Using RACE and circularized RNA RT-PCR techniques, we could isolate and further characterize the complete cDNAs corresponding to the five *Kis2* RNA species (*Kis2*-t1 to -t5) overexpressed in the tumors (Fig. 2.3A). Complete nucleotide sequences of the five cDNAs are accessible in GenBank under accession numbers AY940614 to AY940618.

The five RNA species are mainly nonpolyadenylated, as they are detectable only on Northern blot performed with total RNA. However it is possible to amplify them when performing RT-PCR primed with an oligo(dT), suggesting that at least a little part of these RNAs is polyadenylated. Furthermore, the ESTs from cluster Mm.277876 were obtained from an oligodT-primed library and they contained a 3' end identical to that of those RNAs identified in our 3' RACE study. The canonical AAUAAA polyadenylation signal could not be found in the 3' end of the transcripts. But some genes are known to have alternative signal sequence and we found, at the 3' end, the signal AAAGAA which has been involved in

polyadenylation of some non-AAUAAA-containing mRNAs (72). This signal could represent a putative polyA signal of the *Kis2* gene.

The longest *Kis2* RNA species (t1), which initiates at TSS2 (Transcription Start Site 2), is 2 kb in length and gives rise to species t2, t4, and t5 by alternative splicing (Fig. 2.3A). Splicing sites are canonical and respect the GT/AG rules. The third RNA species (t3) is initiated from a different transcription start site (TSS3), located 0.6 kbp downstream of TSS2. These two TSSs are themselves different from that of EST BB658090 of the Mm.277876 cluster registered in the database, located 0.4 kbp upstream the TSS2 (TSS1) (Fig. 2.3A). Thus, The *Kis2* gene contains at least three different start sites, each one being localized in a CpG island and sufficiently distant from the others suggesting three different promoters.

The coding potential of the five RNA species has been analyzed by computer and we could not find any open reading frame in their sequences. Thus, we concluded that these RNAs belong to the growing number of noncoding RNA group.

The *Kis2* EST unigene cluster (Mm.277876) contains only three ESTs: BB658090 and BB470007, which correspond to the 5' and 3' ends, respectively, of the same cDNA clone isolated from a 12-day embryo eyeball cDNA library, and BB749744, identical to BB470007 and isolated from a pool of tissues. Thus, this gene is very poorly represented in the EST database. In Northern blots and by RT-PCR, the signal corresponding to BB658090 remained undetectable in the *Kis2*-rearranged tumors as in normal tissues. However, we were able to amplify the transcript by nested RT-PCR (data not shown), confirming the rarity of this transcript.

In normal mouse tissues, *Kis2* RNAs were weakly detectable by Northern blot in the thymus only (data not shown), while detection for others tissues was only possible by nested RT-PCR (Fig. 2.3B), suggesting again a very weak transcription of those RNAs in a normal state. Transcripts 4 and 5 are detectable in almost all the tissues tested but with different intensity, the strongest signal being in thymus. Transcripts 1 and 2 are clearly seen in the thymus and spleen only. In general, in our RT-PCR analysis, we found that the intensity of

transcripts 4 and 5 was always higher than that of transcripts 1 and 2. This observation was made in the normal tissue analysis (Fig. 2.3B) but also in the *Kis2*-rearranged tumors (result not shown). However, when the same tumors are analyzed by Northern blot for *Kis2* expression, we observed a stronger signal for the 2 longest transcripts when compared to the 2 smallest (Fig. 2.3A). Thus, it appears that the differences observed in signal intensity between the *Kis2* long and small transcripts in Fig. 2.3B is probably the result of a skewed PCR in favor of the smaller transcripts since the same pair of oligonucleotides was used for transcript 1, 2, 4, and 5 amplification. Specific detection of transcript 3 in normal tissues was not possible by RT-PCR because its sequence is entirely included in that of transcripts 1 and 2 (Fig. 2.3A).

Until now, no orthologues of the *Kis2* gene have been found, as it is absent from EST database of other species. Interestingly, a comparison of genomic DNA between mouse, rat, and human revealed some highly conserved regions. In particular, striking sequence conservations are observed in the six regions corresponding to TSSs, splice sites, and a 250-bp sequence located in the last quarter of the *Kis2* transcription unit (Fig. 2.4, region 5). This suggests that the *Kis2* gene might be present in other species such as rat and human, even though ESTs have not been reported yet.

#### **2.4.4 *Phf6* is also overexpressed in *Kis2*-rearranged tumors.**

The *Phf6* gene is located about 250 kbp away from the RadLV integrations (Fig. 2.2B). Nevertheless, we verified the expression of all the genes and ESTs around the integration site. To our surprise, we observed an important overexpression of this gene specifically in *Kis2*-rearranged tumors (Fig. 2.5A). Tumors with strong overexpression of *Phf6* (T1, T2, T3, T5, and T6) were also those which strongly expressed *Kis2* RNAs, and they corresponded to the most clonal *Kis2*-rearranged tumors. Tumor T4 was partially rearranged for *Kis2* since the rearrangement was detected by PCR only. Interestingly, it expressed intermediate levels of *Phf6* (Fig. 2.5A). Two tumors showed barely detectable levels of *Phf6* although we were unable to detect a provirus in the 6-kbp region that

encompass the *Kis2* integration locus. However, we cannot exclude the possibility that a retroviral integration has occurred outside this region.

The mouse *Phf6* gene has 11 exons and encodes a 365-amino-acid protein, which contains two Plant Homology Domain (PHD) fingers. *Phf6* was unknown until it was recently implicated in the Börjeson-Forssman-Lehmann X-linked mental retardation syndrome (BFLS; OMIM 301900) in humans (43). To isolate complete cDNAs of the *Phf6* gene, we constructed a cDNA library from a rearranged tumor and we obtained clones corresponding to three different transcripts (labeled t2, t3, and t7 in Fig. 2.5B and 2.6).

The longest (*Phf6*-t3) has a 4-kb 3' untranslated region and is the most abundant in our tumors. This is also the major transcript found in the database as it constitutes 96% of ESTs in *Phf6* cluster Mm.26870 (400 total ESTs). The ESTs corresponding to this transcript are described in a broad variety of embryonic and adult tissues, with reports of its presence in brain and activated spleen being the most frequent.

The second transcript in size (*Phf6*-t2) also has 11 exons, but lacks the long 3' untranslated region (Fig. 2.6). It is weakly expressed in our tumors, as only a very weak signal is observed on poly(A) RNA Northern blots (not shown). Furthermore, it represents only 2% of the ESTs present in the Mm.26870 cluster (see also the legend to Fig. 2.6). Both *Phf6*-t2 and *Phf6*-t3 encode an identical 365 amino acid protein. We could detect expression of t2 and t3 by nested RT-PCR (Fig. 2.5B) in almost all the normal tissues, with the weakest expression being in muscle. The primers used in this assay did not allow us to distinguish *Phf6* transcript 2 from transcript 3. However, we obtained identical results with primers specific to transcript 3 only (reverse primer located at the 3' end of the 3' untranslated region; data not shown). Thus, the signal observed in Fig. 2.5B corresponds almost entirely to transcript 3.

Interestingly, we also found a third smaller transcript (*Phf6*-t7) in our cDNA library, lacking exons 2 to 5 inclusively (Fig. 2.6). This species has never been described before, and it gives rise to a protein lacking the first PHD domain. Its detection required nested RT-PCR and it seemed to be expressed mainly in normal thymus (Fig. 2.5B).

Analysis of all the ESTs from the Phf6 cluster revealed four other alternative transcripts in the mouse, although they are rarely represented in the database (Phf6-t1, Phf6-t4, Phf6-t5, and Phf6-t6, Fig. 2.6). Phf6-t1 is composed of exon 1 and a part of intron 1. It contains an open reading frame totally different from that of PHF6, since the normal ATG start codon is in exon 2. Phf6-t4 lacks exons 2 and 3 and codes for a protein that still contains the two plant homology domains but has lost the first 80 amino acids at the N-terminal end. This transcript was detected in thymus, lungs, ovaries, kidneys, liver, and spleen (Fig. 2.5B). Phf6-t5 has lost exons 6 to 10 and gives rise to a protein lacking the second PHD domain. Phf6-t6 lacks exons 9 and 10, resulting in a protein 41 amino acids shorter at the C-terminal end that has conserved the two PHD domains.

In humans, the PHF6 protein localizes to the nucleolus, and each of the four nuclear localization signals (NLSs) identified seems to be important (43) (Fig. 2.6). Using enhanced green fluorescent protein (EGFP) fusion protein constructs, we found the same localization in mouse NIH 3T3 cells. Protein from Phf6-t2 did localize to the nucleolus like the human homologue (Fig. 2.7). Interestingly, protein from transcript 7 of Phf6 (labeled PHF6 t7 on Fig. 2.7), which lost all four NLSs, still localized to the nucleus but no longer to the nucleolus (Fig. 2.7, middle panel). These results were the same whether the EGFP was in the N- or C-terminal position in the fusion protein constructs. As NLSs are absent from the PHF6 t7 protein, this suggests that other unidentified NLSs are present in the protein or that the protein interacts with other nuclear proteins to be imported.

## 2.5 DISCUSSION

In this article, we report the identification of *Kis2*, a novel RadLV common integration site on the X chromosome. Before the mouse genome sequencing was completed, searching for genes located in a region was painstaking and time-consuming, and as a consequence, common integration sites were often associated with a unique gene. We analyzed the expression of all the genes present at the *Kis2* locus and, surprisingly, we found that two genes were overexpressed in the rearranged tumors: first, *Kis2*, which is located close to the retrovirus integration site, and second, the *Phf6* gene, situated about 250 kbp away. Thus, this article is also the first report on a retroviral integration that affects the expression of two different distant genes. Indeed, it is now possible to investigate the effect of an integrated retrovirus on all the genes from a given region since almost the entire mouse genome is now available and well annotated.

First of all, our results show that genes situated quite far from the integration site could be activated as well. This has been previously described in the case of Moloney murine leukemia virus when integrated at the *Mlvi4* locus, 300 kbp from the *c-myc* locus (37). This fact was rarely described in the literature until recently, mainly due to the lack of the appropriate genetic information. Recent high-throughput studies using the mouse genome sequence database have reported common retroviral integrations in regions located much further away from the target gene (11, 32, 68). It is clear that identification of target genes located further away from the common retroviral integration sites will now be possible.

How can retroviral integrations in regions located far away from a target gene act so efficiently on its expression? Specific chromatin structures could be involved so that distant regions are brought closer. Alternatively, key regulatory elements of gene expression could be present at the integration site and be affected by the presence of the retrovirus. So retroviruses could be tools not only to find oncogenes, but also to find their regulatory elements, as already mentioned by Kim et al. (32).

It is also interesting that not all the genes in the region are affected by the integration. This had already been reported in the case of SL3-3 when integrated in the *Ras* locus: only the expression of *Ras* is affected, whereas the expression of *unr* is not, despite its localization close to the integration site (48). This could result from the fact that some genes are located in an active area of the chromatin while others are not. It can also be illustrated by Moloney murine leukemia virus integrations at the *XPCL1* locus (Fig. 2.2B) which affect the expression of EST AI464896 (27), whereas RadLV did not. A response of a given promoter to a specific enhancer could also be involved so that different enhancers would affect gene expression differently. Indeed, long terminal repeat sequences have a great influence on the insertional activation, as recently shown by Johnson et al. (29). In light of this, it would be interesting to know if Moloney murine leukemia virus or SL3-3 enhancers do affect the expression of *Kis2* and *Phf6* when integrated at the *Kis2* locus. Also relevant to this question, it is interesting that antisense retroviral integrations affect transcription from TSS2 and TSS3 in a similar fashion, whereas integrations in the sense orientation affect expression from TSS3 to a much greater extent (Fig. 2.3A).

*Kis2* is a novel gene that has never been described before. It is poorly represented in the mouse database. Indeed, we found that *Kis2* is weakly expressed in all the normal tissues, the thymus being the highest expressor (Fig. 2.3B). To date, no ESTs specific to *Kis2* have been described in other species. However, the *Kis2* transcription unit, or at least some of its features, seems to be conserved in rat and human, as alignment of the genomic regions reveals highly conserved areas in the sequence overlapping TSSs and splice sites.

The main characteristic of *Kis2* transcripts is that they contain only minute open reading frames, thus classifying them as noncoding RNAs. A few years ago, noncoding RNAs were considered noise in the transcriptional system. But recent studies from the human and mouse genome sequencing projects revealed that a real noncoding RNA dimension remains to be explored. The function of most of those noncoding RNAs is still unknown. However, a few of them have now been functionally characterized. Noncoding RNAs are implicated in mechanisms such as imprinting (H19) (57), dosage compensation transcription



regulation (*Xist* [6], *Tsix* [38], and *roX* [49]), (*SRA* [36] and *7SK* [33]), regulation of transcript translation by antisense effect involving a particular micro-RNA (*lin-4* [54] and *let-7* [62]), and finally oxidative stress (*adapt33* [73] and *gadd45* [25]).

The function of *Kis2* remains to be determined. One hypothesis about the role of noncoding transcripts argues against a function of the RNA itself but rather stipulates that only the transcription across the locus would be important, as this is the case for the  $\beta$ -globin locus control region (16). However, if this is the case for the *Kis2* locus, it is difficult to understand the meaning of the three promoters and the six alternative transcripts. Such characteristics have been reported for others noncoding RNAs such as *Xist* and *Tsix*, for which no functions for the spliced species have been attributed yet (52, 53, 66). In the case of noncoding RNAs, the sequence could really be important either at the level of antisense functions or as a structure. Because of their sequence differences, the *Kis2* RNA species very likely have distinct properties, which argues in favor of a function for the *Kis2* RNA per se.

Another interesting particularity of *Kis2* RNAs is that they apparently lack a poly(A) tail, as we could not detect them on Northern blot performed with poly(A) RNA. The poly(A) tail is of great importance for mRNA, both at the level of transcription regulation and for stability. Sequence elements promoting mRNA instability have been described, such as those named AU-rich elements, which are located in the 3' untranslated region of mRNA and control their deadenylation (75). Thus, some RNAs could be present as polyadenylated or deadenylated forms as a result of poly(A) tail length regulation, and are qualified as bimorphic RNAs. Alternatively, some mRNAs have sequences in their 3'untranslated region named poly(A)-limiting elements, which limit the length of the poly(A) tail (17, 19). Thus, those RNAs are generated with a poly(A) tail containing less than 20 A's, which leads eventually to transcript instability in response to some regulatory factors. The same situation could occur for the *Kis2* mRNAs: They could have a poly(A) tail too short to be successfully purified by oligo(dT) affinity chromatography, but still long enough to be amplified by oligo(dT)-primed RT-PCR. Such poly(A)-limiting element consensus sequences are not present in the *Kis2* 3' region, although other sequences could be responsible for limiting the poly(A) length, as this is also the case for immunoglobulin M mRNA (58). Finally, the lack

of a poly(A) tail of normal length in *Kis2* RNAs could be in part responsible for their lower representation in the databases since most cDNA libraries are primed by oligo(dT).

The *Phf6* gene was discovered recently in human and was implicated in the Börjeson-Forssman-Lehmann syndrome (43). BFLS is characterized by moderate to severe mental retardation, epilepsy, hypogonadism, hypometabolism, obesity with marked gynecomastia, swelling of subcutaneous tissue of the face, narrow palpebral fissures, and large but not deformed ears (5). The *Phf6* gene encodes a 365-amino-acid protein, which has two plant homology domain fingers. Several PHF6 mutations associated with BFLS have been found (43): some are missense mutations, with one of them reported to affect a cysteine in the first plant homology domain, and others are truncation mutations. One of those truncation mutations occurs in the beginning of the protein, resulting in its complete loss, while another one, which appears to be recurrent in BFLS (44), occurs at the end of the protein (R342X).

The multiple symptoms which characterize this syndrome can be correlated with the expression of the gene in a broad variety of tissues, and with the existence of several alternatively spliced transcripts. Although Phf6 expression is detected in most human and mouse tissues, expression is low and only detected by nested RT-PCR. However, in the *Kis2*-rearranged tumors, its expression was easily detectable by Northern blot and RT-PCR analysis (not shown and Fig. 2.5). In consequence, this article is the first report suggesting an involvement of *Phf6* in the development of leukemia.

PHD finger composition is close to the RING finger and LIM domain (1), which both function in association with two  $Zn^{2+}$  atoms. Although the PHD finger function is still unknown, it is mainly found in proteins involved in transcriptional regulation (1). Two groups can be distinguished: the first one including transcriptional activators, repressors and cofactors (28), and the second one including proteins of chromatin modulating complexes such as acetyltransferases or acetyltransferase-containing complexes such as p300 or CBP (4). To corroborate this, the PHD finger of the transcription factor SPBP is involved in chromatin-mediated transcriptional regulation as a protein-protein interaction domain (46). Furthermore, naturally occurring single amino acid substitutions in the PHD finger of the

ATRX and AIRE1 proteins predispose individuals to  $\alpha$ -thalassemia, developmental defects and autoimmune diseases (3, 14). Interestingly, ALL1, the human homologue of *Drosophila* trithorax, and other PHD finger-containing genes are also frequently disrupted by chromosomal rearrangements that occur in acute lymphoid and myeloid leukemias (8, 18, 51, 56, 69).

As in human, the mouse PHF6 protein was found to localize to the nucleolus. This organelle is mainly known as a site dedicated to ribosome biogenesis. Other events take place in the nucleolus, such as for example, the maturation of polymerase III transcripts (7). The nucleolus also seems to behave as a sequestering compartment for regulatory complexes. For example, the p19ARF tumor suppressor, which is implicated in p53 regulation, exclusively localizes to the nucleolus. Cytoplasmic p53 degradation is mediated by the oncoprotein Mdm2 and p19ARF acts by complexing and sequestering mdm2 in the nucleolus, thus preventing its nuclear export and p53 degradation (55). Nucleolus also hosts viral protein, such as HIV-Rev (26). According to these observations, the nucleolus still appears as a “dark box”, now including Phf6.

However, it is likely that *Phf6* serves fundamental functions in the cell since it is highly conserved among species. Several transcripts are present in the database, but the major one is transcript 3 (Fig. 2.6). It differs from the others by its long 3' untranslated region, which probably serves some regulatory or stability goals. Transcripts 4 and 6 give rise to a protein which retains the two PHD fingers, while the protein encoded by transcript 5 has lost the second one. Interestingly, we found another alternative transcript (transcript 7, Fig. 2.6) which has never been described before. It appears very weakly expressed, mainly in the thymus. This splicing eliminates exons 2 to 5, giving rise to a protein initiated at an ATG codon further downstream and thus lacks the first PHD domain and all the identified nuclear/nucleolar localization signals. Furthermore, the corresponding protein no longer localizes to the nucleolus, but still remains in the nucleus.

Finally, it is tempting to make a link between the *Kis2* and *Phf6* genes which are both expressed simultaneously in *Kis2*-rearranged tumors. Because of the noncoding nature of

*Kis2* RNAs, one could imagine a trans-action of these RNAs on the *Phf6* locus, which would lead to its overexpression. Otherwise, because of the relative proximity of *Kis2* and *Phf6* on the chromatin, a cis-action of the transcription across the *Kis2* locus on the *Phf6* expression, mainly through a chromatin structure remodeling could be imagined.

There is no doubt that the *Kis2* locus is involved in thymomas occurrence since it is targeted by three different retroviruses that induce T-cell leukemia: Moloney, SL3-3 and RadLV. Since there is some variability in tumor clonality, some being almost clonal and some not at all, it is tempting to speculate that the role of the *Kis2* locus is more oriented towards tumor progression rather than tumor initiation. We must now further characterize *Kis2* RNAs and PHF6 protein and determine if either the *Kis2* or *Phf6* gene or both have oncogenic potential.

### **Acknowledgments**

We thank Elsy Edouard and Laurent Poliquin for helpful discussions. This work was supported by grant MOP-37994 from the Canadian Institutes of Health Research.

## 2.6 REFERENCES

1. Aasland, R., T.J Gibson, and A.F Stewart. 1995. The PHD finger : implication for chromatin-mediated transcriptional regulation. *Trends Biochem. Sci.* 20 : 56-59.
2. Ben-David, Y., E. Yefenof, and M. Kotler. 1987. Clonal analysis of radiation leukemia virus-induced leukemic and preleukemic murine cells. *Cancer Res.* 47:6590-6594.
3. Björjes P., M. Halonen, J.J. Palvimo, M. Kolmer, J. Aaltonen, P. Ellonen, J. Perheentupa, I. Ulmanen, and L. Peltonen. 2000. Mutations in the AIRE gene: effects on subcellular location and transactivation function of the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy protein. *Am. J. Hum. Genet.* 66 : 378-392.
4. Bordoli, L., S. Husser, U. Luthi, M. Netsch, H. Osmani, and R. Eckner. 1992. Functional analysis of the p300 acetyltransferase domain : the PHD finger of p300 but not of CBP is dispensable for enzymatic activity. *Nucleic Acids Res.* 29 : 4462-4471.
5. Börjeson, M., H. Forssman, O. & Lehmann. 1962. An X-linked, recessively inherited syndrome characterized by grave mental deficiency, epilepsy, and endocrine disorder. *Acta Med. Scand.* 171 : 13-21.
6. Brockdorff, N., A. Ashworth, G.F. Kay, V.M. McCabe, D.P. Norris, P.J. Cooper, S. Swift, and S. Rastan. 1992. The product of the mouse *Xist* gene is a 15 Kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* 71 : 515-526.
7. Carmo-Foncesca, M., L. Mendes-soares and I. Campos. 2000. To be or not to be in the nucleolus. *Nat. Cell. Bio.* 2 : 107-112.
8. Chesi, M., E. Nardini, R.S. Lim, K.D. Smith, W.M. Kuehl, and P.L. Bergsagel. 1998. The t(4;14) translocation in myeloma dysregulates both FGFR3 and a novel gene, MMSET, resulting in IgH/MMSET hybrid transcripts. *Blood* 92 : 3025-3034.

9. Corcoran, L.M., J.M. Adams, A.R. Dunn, and S. Cory. 1984. Murine T lymphomas in which the cellular myc oncogene has been activated by retroviral insertion. *Cell* 37 : 113-122.
10. Cuypers, H.T., G. Selten, W. Quint, M. Zijlstra, E.R. Maandag, W. Boelens, P. van Wezenbeek, C. Melief, and A. Berns. 1984. Murine leukemia virus-induced T-cell lymphomagenesis: integration of proviruses in a distinct chromosomal region. *Cell* 37 : 141-150.
11. Denicourt, C., C.A. Kozak, and E. Rassart. 2003. Grisl, a new common integration site in Graffi murine leukemia virus-induced leukemias: overexpression of a truncated cyclin D2 due to alternative splicing. *J. Virol.* 77 : 37-44.
12. Dudley, J.P., J.A Mertz, , M. Lozano, and Dr Broussard. 2002. What retroviruses teach us about the involvement of c-Myc in leukemias and lymphomas. *Leukemia* 16 : 1086-1098.
13. Filmus, J., and S.B. Selleck. 2001. Glypicans : proteoglycans with a surprise. *J. Clinic. Investig.* 108 : 497-501.
14. Gibbons, R.J., S. Bachoo, D.J. Picketts, S. Aftimos, B. Asenbauer, J. Bergoffen, S.A. Berry, N. Dahl, A. Fryer, K. Keppler, K. Kurosawa, M.L. Levin, M. Masuno, G. Neri, M.E. Pierpont, S.F. Slaney, and D.R. Higgs. 1997. Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain. *Nat.Genet.* 17 : 146-148.
15. Girard, L., Z. Hanna, N. Beaulieu, C.D. Hoemann, C. Simard, C.A. Kozak, and P. Jolicoeur. 1996. Frequent provirus insertional mutagenesis of Notch1 in thymomas of MMTVD/myc transgenic mice suggests a collaboration of c-myc and Notch1 for oncogenesis. *Genes Dev.* 10 : 1930-1944.
16. Gribnaud, J., K. Diderich, S. Pruzina, R. Calzolari, and P. Frascati. 2000. Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. *Mol. Cell* 5 : 377-386.
17. Gu, H., J.D. Gupta, and D.R. Schoenberg. 1999. The poly(A)-limiting element is a conserved cis-acting sequence that regulates poly(A) tail length on nuclear pre-mRNAs. *Proc. Natl. Acad. Sci. USA.* 96 : 8943-8948.

18. **Gu, Y., T. Nakamura, H. Alder, R. Prasad, O. Canaani, G. Cimino, C.M. Croce, and E. Canaani.** 1992. The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to *Drosophila trithorax*, to the AF-4 gene. *Cell* 71 : 701-708.
19. **Gu, H., and D.R. Schoenberg.** 2003. U2AF modulates poly(A) length control by the poly(A)-limiting element. *Nucleic Acids Res.*, 31 : 6264-6471.
20. **Hadland, B.K., S.S. Huppert, J. Kanungo, Y. Xue, R. Jiang, T. Gridley, R.A. Conlon, A.M. Cheng, R. Kopan, and G.D. Longmore.** 2004. A requirement for Notch1 distinguishes 2 phases of definitive hematopoiesis during development. *Blood*. 104 : 3097-3105.
21. **Hahn, W.C., and R.A. Weinberg.** 2002. Modelling the molecular circuitry of cancer. *Nat. Reviews*. 2 : 331-341.
22. **Hanahan, D., and R.A. Weinberg.** 2000. The hallmarks of cancer. *Cell* 100 : 57-70
23. **Hanna, Z., M. Janowski, P. Tremblay, X. Jiang, A. Milatovich, U. Francke, and P. Jolicoeur.** 1993. The Vin-1 gene, identified by provirus insertional mutagenesis, is the cyclin D2. *Oncogene* 8 : 1661-1666.
24. **Hoemann, C.D, N. Beaulieu, L.Girard, N. Rebai and P. Jolicoeur.** 2000. Two distinct *Notch1* mutant alleles are involved in the induction of T-cell leukemia in c-myc transgenic mice. *Mol. Cell. Biol.* 20 : 3831-3842.
25. **Hollander, M.C., I. Alamo, and A.J. Fornace Jr.** 1996. A novel DNA damage-inducible transcript, gadd45, inhibits cell growth, but lacks a protein product. *Nucleic Acids Res.* 24 : 1589-1593.
26. **Hope, T.J.** 1999. The ins and outs of HIV Rev. *Arch. Biochem. Biophys.* 365 : 186-191.
27. **Hwang, H.C., C.P. Martins, Y. Bronkhorst, E. Randel, A. Berns, M. Fero and B.E. Clurman.** 2002. Identification of oncogenes collaborating with p27Kip1 loss by insertional mutagenesis and high-throughput insertion site analysis. *Proc. Natl. Acad. Sci. U S A.* 99 : 11293-11298.
28. **Ikegawa, S., M. Isomura, Y. Koshizuka, and Y. Nakamura.** 1999. Cloning and characterisation of ASH2L and Ash21, human and mouse homologs of the *drosophila ash2* gene. *Cytogenet. Cell. Genet.* 84 : 167-172.

29. **Johnson, C., P.A. Lobelle-Rich, A. Puetter, and L.S. Levy.** 2005. Substitution of feline leukemia virus long terminal repeat sequences into murine leukemia virus alters the pattern of insertional activation and identifies new common insertion sites. *J Virol.* 79 : 57-66.
30. **Joosten, M., Y. Vankan-Berkhoudt, M. Tas, M. Lunghi, Y. Jenniskens, E. Parganas, P.J.M. Valk, E. Van den Akker, and R. Delwel.** 2002. Large-scale identification of novel potential disease loci in mouse leukemia applying an improved strategy for cloning common virus integration site. *Oncogene* 21 : 7247-7255.
31. **Kaplan, H.S.** 1967. On the natural history of the murine leukemias: presidential address. *Cancer Res.* 27 : 1325-1340
32. **Kim, R., A. Trubetskoy, T. Suzuki, N.A. Jenkins, N.G. Copeland, and J. Lenz.** 2003. Genome based identification of cancer genes by proviral tagging in mouse retrovirus-induced T-cell lymphomas. *J. Virol.* 77 : 2056-2062.
33. **Krause, M.O.** 1996. Chromatin structure and function: the heretical path to an RNA transcription factor. *Biochem. Cell. Biol.* 74 : 623-632.
34. **Kuhn, J., and S. Binder.** 2002. RT-PCR analysis of 5' and 3'-end-ligated mRNAs identifies the extremities of *cox2* transcripts in pea mitochondria. *Nucleic Acids Res.* 30 : 439-446.
35. **Lambert, J., D. Bergeron, C.A. Kozak, and E. Rassart.** 1999. Identification of a common site of provirus integration in radiation leukemia virus-induced T-cell lymphomas in mice. *Virology.* 264 : 181-186.
36. **Lanz, R.B., N.J. McKenna, S.A. Onate, U. Albrecht, J. Wong, S.Y. Tsai, M.J. Tsai, and B.W. O'Malley.** 1999. A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* 97 : 17-27.
37. **Lazo, P.A., J.S. Lee, and P.N. Tsichlis.** 1990. Long-distance activation of the *Myc* protooncogene by provirus insertion in *MLV*-1 or *MLV*-4 in rat T-cell lymphomas. *Proc. Natl. Acad. Sci. USA.* 87 : 170-173.
38. **Lee, J.T., S.L. Davidow and D. Warshawsky.** 1999. *Tsix*, a gene antisens to *Xist* at the X-inactivation center. *Nat. Genet.* 21 : 400-404.



39. **Li, J., H. Shen, K.L. Himmel, A.J. Dupuy, D.A. Largaespada, T. Nakamura, J.D. Jr Shaughnessy, N.A. Jenkins, and N.G. Copeland.** 1999. Leukaemia disease genes: large-scale cloning and pathway predictions. *Nat. Genet.* 23 : 348-353.
40. **Lieberman, M., A. Declève, E.P. Gellmann, and H.S. Kaplan.** 1977. Biological and serological characterization of the C-type RNA viruses isolated from C57BL/Ka strain of mice.II. Induction and propagation of the isolates in radiation-induced leukemogenesis and related viruses, p.231-246. In J.F. Duplan (ed.), INSERM Symposium no.4, Paris. France.
41. **Lieberman, M., A. Declève, P. Ricciardi-Castagnoli, J. Bonier, O.J. Finn, and H.S. Kaplan.** 1979. Establishment, characterization and virus expression of cell lines derived from radiation-and virus-induced lymphomas of C57BL/Ka mice. *Int. J. Cancer* 24 : 168-177.
42. **Lieberman, M., O. Niwa, A. Declève, and H.S. Kaplan.** 1973. Continuous propagation of radiation leukemia virus on a C57BL mouse embryo fibroblast line with attenuation of leukemogenic activity *Proc. Natl. Acad. Sci. USA* 70 : 1250-1253
43. **Lower, K.M., G. Turner, B.A. Kerr, K.D. Mathews, M.A. Shaw, A.K. Gedeon, S. Schelley, H.E. Hoyme, S.M. White, M.B. Delatycki, A.K. Lampe, J. Clayton-Smith, H. Stewart, C.M. van Ravenswaay, B.B. de Vries, B. Cox, M. Grompe, S. Ross, P. Thomas, J.C. Mulley, and J.Gecz.** 2002. Mutations in PHF6 are associated with Börjeson-Forssman-Lehmannsyndrome. *Nat. Genet.* 32 : 661-665.
44. **Lower, K.M., G. Solders, M.L. Bondeson, J. Nelson, A. Brun, J. Crawford, G. Malm, M. Borjeson, G. Turner, M. Partington, and J. Gecz.** 2004. 1024C>T (R342X) is a recurrent PHF6 mutation also found in the original Börjeson-Forssman-Lehmann syndrome family. *Eur. J. Hum. Genet.* 12 : 787-789.
45. **Lund, A.H., G. Turner, A. Trubetskoy, E. Verhoeven, E. Wientjens, D. Hulsman, R. Russel, R.A. DePinho, J. Lenz, and M. Van Lohuizen.** 2002.

Genome wide retroviral insertional tagging of genes involved in cancer in Cdkn2a-deficient mice. *Nat. Genet.* 32 : 160-165

46. **Lyngso, C., G. Bouteiller, C.K. Damgaard, D. Ryom, S. Sanchez-Munoz, P.L. Norby, B.J. Bonven, and P. Jorgensen.** 2000. Interaction between the transcription factor SPBP and the positive cofactor RNF4 : an interplay between protein binding zing fingers. *J. Biol. Chem.* 275 : 26144-26149.
47. **Maillard, I., A.P. Weng, A.C. Carpenter, C.G. Rodriguez, H. Sai, L. Xu, D. Allman, J.C. Aster, and W.S. Pear.** 2004. Mastermind critically regulates Notch-mediated lymphoid cell fate decisions. *Blood.* 104 : 1696-1702.
48. **Martin-Hernandez, J., A.B. Sorensen, and F.S. Pedersen.** 2001. Murine leukemia virus proviral insertions between the N-ras and unr genes in B-cell lymphoma DNA affect the expression of N-ras only. *J. Virol.* 75 : 11907-11912.
49. **Meller, V.H., K.H. Wu, G. Roman, M.I. Kuroda, and R.L. Davis.** 1997. roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system. *Cell* 88 : 445-457.
50. **Mikkers, H., J. Allen, P. Knipscheer, L. Romeyn, A. Hart, E. Vink, and A. Berns.** 2002. High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. *Nat. Genet.* 32 : 153-159.
51. **Nakamura, T., H. Alder, Y. Gu, R. Prasad, O. Canaani, N. Kamada, R.P. Gale, B. Lange, W.M. Crist, P.C. Nowell, et al.** 1993. Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leukemia share sequence homology and/or common motifs. *Proc Natl Acad Sci U S A.* 90 : 4631-4635.
52. **Nesterova, T.B., C.M. Johnston, R. Appanah, A.E. Newall, J. Godwin, M. Alexiou, and Brockdorff N.** 2003. Skewing X chromosome choice by modulating sense transcription across the Xist locus. *Genes Dev.* 17 : 2177-2190.
53. **Ogawa, Y., J.T. Lee.** 2003. Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. *Mol. Cell* 11 : 731-743.

54. Olsen, P.H., and V. Ambros. 1999. The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* 216 : 671-680.
55. Olson, M.O., M. Dundr, and A. Szebeni. 2000. The nucleolus: an old factory with unexpected capabilities. *Trends Cell. Biol.* 10 : 189-196.
56. Parry, P., M. Djabali, M. Bower, J. Khristich, M. Waterman, et al. 1993. Structure and expression of the human trithorax-like gene 1 involved in acute leukemia. *Proc Natl. Acad. Sci. USA.* 90 : 4738-4742.
57. Peters, J. 2000. Imprinting: silently crossing the boundary. *Genome Biol.* 1 : reviews 1028.
58. Phillips, C., S. Jung, and S.I. Gunderson. 2001. Regulation of nuclear poly(A) addition controls the expression of immunoglobulin M secretory mRNA. *EMBO J.* 20 : 6443-6452.
59. Poliquin, L., D. Bergeron, J.L. Fortier, Y. Paquette, R. Bergeron, and E. Rassart. 1992. Determinants of thymotropism in Kaplan radiation leukemia virus and nucleotide sequence of its envelope region. *J. Virol.* 66:5141-5146
60. Rassart, E., P. Sankar-Mistry, G. Lemay, L. Desgroseillers, and P. Jolicoeur. 1983. New class of leukemogenic ecotropic recombinant murine leukemia virus isolated from radiation induced thymomas of C57BL/Ka mice. *J. Virol.* 45 : 565-575.
61. Rassart, E., M. Shang, Y. Boie, and P. Jolicoeur. 1986. Studies on emerging Radiation leukemia virus variants in C57BL/Ka mice. *J. Virol.* 58 : 96-106.
62. Reinhart, B.J., F.J. Slack, M. Basson, A.E. Pasquinelli, J.C. Bettinger, A.E. Rougvie, H.R. Horvitz, and G. Ruvkun. 2000. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, 403 : 901-906.
63. Sambrook, J., and D.W Russell. 2001. molecular cloning : a laboratory manual, 3<sup>rd</sup> ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
64. Selten, G., H.T. Cuypers and A. Berns. 1985. Proviral activation of the putative oncogene *Pim-1* in MuLV induced T-cell lymphomas. *EMBO J.* 4 : 1793-1798.

65. Selten, G., H.T. Cuypers, M. Zijlstra, C. Melief and A. Berns. 1984. Involvement of c-Myc in MuLV-induced T cell lymphomas in mice : frequency and mechanisms of activation. *EMBO* 3 : 3215-3222.
66. Shibata, S. and J.T. Lee. 2003. Characterisation and quantitation of differential Tsix transcripts : implication for Tsix function. *Hum. Mol. Gen.* 12 : 125-136.
67. Sørensen, A.B., M. Duch, H.W. Amtoft, P. Jørgensen, and F.S. Pedersen. 1996. Sequence Tags of provirus integration sites in DNAs of tumors induced by the murine retrovirus SL3-3. *J. Virol.* 70 : 4063-4070.
68. Suzuki, T., H. Shen, K. Akabi, H.C. Morse III, J.D. Malley, D.Q. Naiman, N.A. Jenkins, and N.G. Copeland. 2002. New genes involved in cancer identified by retroviral tagging. *Nat. Genet.* 32 : 166-174.
69. Tkachuk, D.C., S. Kohler, and M.L. Cleary. 1992. Involvement of a homologue of *Drosophila trithorax* by 11q23 chromosomal translocations in acute leukemias. *Cell* 71 : 691-700.
70. Tremblay, P.J, C.A. Kosak, and P. Jolicoeur. 1992. Identification of a novel gene, Vin-1, in murine leukemia virus induced T-cell leukemias by provirus insertional mutagenesis. *J. Virol*, 66 : 1344-1353.
71. Van Lohuizen, M., S. Verbeek, P. Krimperfort, J. Domen, C. Saris, T. Radaszkiewicz, and A. Berns. 1989. Predisposition to lymphomagenesis in pim-1 transgenic mice: cooperation with c-myc and N-myc in murine leukemia virus-induced tumors. *Cell* 56 : 673-682.
72. Wallace, A.M., D. Brinda, S.E. Ravnik, V. Tonk, N.A. Jenkins, D.J. Gilbert, N.G. Copeland, and C.C. MacDonald. 1999. Two distinct forms of the 64,000 Mr protein of the cleavage stimulation factor are expressed in mouse male germ cells. *Proc. Natl. Acad. Sci. USA.* 96 : 6763-6768.
73. Wang, Y., D.R. Crawford, and K.J.A. Davies. 1996. *adapt33*, a novel oxidant-inducible RNA from hamster HA-1 cells. *Arch. Bioch. Bioph.* 332 : 255-260.
74. Weng, A.P., A.A. Ferrando, W. Lee, J.P. 4th Morris, L.B. Silverman, C. Sanchez-Irizarry, S.C. Blackdow, A.T. Look, and J.C. Aster. 2004.

Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science*. 306 : 269-271.

75. **Xu, N., C.Y. Chen, and A.B. Shyu.** 1997. Modulation of the fate of cytoplasmic mRNA by AU-rich elements: key sequence features controlling mRNA deadenylation and decay. *Mol Cell Biol*. 17 : 4611-4621.
76. **Yanagawa S, Lee JS, Kakimi K, Matsuda Y, Honjo T, and A. Ishimoto.** 2000. Identification of Notch1 as a frequent target for provirus insertional mutagenesis in T-cell lymphomas induced by leukemogenic mutants of mouse mammary tumor virus. *J. Virol*. 74 : 9786-9791.

## 2.7 FIGURES LEGEND

**Figure 2.1 :** Positions of RadLV integrations in *c-myc*, *Pim1* and *Notch1* loci. The probes used are represented by dashed boxes. Arrows above the maps indicate the positions and the orientation of the RadLV integrations. Tumor 140.2 was rearranged for both *c-myc* and *Pim1*. Dark boxes represent the exons. B, BamHI; H, HindIII; K, KpnI; P, PstI; S, SacII, X, XbaI.

**Figure 2.2 :** Southern blot analysis of some RadLV-induced leukemias. (A) Control tissue (C) and tumor DNAs (R) digested with restriction enzymes as indicated and hybridized with *Kis2* probe. Note the absence of normal allele in rearranged tumor 2, corresponding to a male mouse. (B). Genomic map of the *Kis2* locus. The RadLV integrations and orientations are represented by arrows (T1 to T7). SL3 and Moloney (*XPCL1*) integrations are also represented. Genes and EST clusters of the region are schematized by grey boxes. The arrow in the box indicates the transcription orientation.

**Figure 2.3 :** (A) Northern blot analysis of the *Kis2* RNA species and schematic representation of their structure. Left side : total RNA of rearranged and non rearranged (NR) tumor was hybridized with *Kis2* RTB3 DNA probe (indicated in the right panel). T1, T2, T3 correspond to tumors with an antisense retroviral integration, and T5, T6 to tumors with a sense retroviral integration. BLVL3 shows RNA extracted from the BLVL3 tumor cell line. Right side : schematic representation of the five RNA species (t1 to t5) as deduced from sequencing. The splicing events are indicated. TSS : Transcription Start Site. The *Kis2* EST BB658090 is also shown. (B) Semi-quantitative analysis by nested RT-PCR of the expression level of *Kis2* transcripts in normal mouse tissues. The actin gene was used as an internal standard.

**Figure 2.4 :** Analysis of the *Kis2* sequence conservation among species. The most conserved regions, identified 1 to 6 on the top schematic representation were aligned for rat, human, and mouse. The 3 transcription start sites (TSS) and the splice donor (SD) and acceptor sites (SA) are indicated above the sequence and the nucleotide underlined. Some

sequences have been deleted (//) to simplify.

**Figure 2.5 :** Analysis of *Phf6* gene expression in tumors and normal mouse tissues. (A) *Kis2*-rearranged (T1 to T6) and non-rearranged (NR) tumors by RT-PCR. (B) Nested RT-PCR on a panel of normal tissues. Primers used allowed amplification of transcript 2 and 3, as well as transcript 4 and 7, according to the expected sizes (800 pb for transcript 2 and 3, 550 pb for transcript 4, and 340 pb for transcript 7). Two tumors (\*) were non-rearranged for *Kis2* but expressed low levels of *Phf6*. The actin gene was used as an internal standart.

**Figure 2.6.** Schematic representation of all the *Phf6* transcripts. All the ESTs (400 total) of the *Phf6* Unigen cluster Mm.26870 were analysed using the NCBI and Ensembl Blast programs. The relative abundance of each transcript is indicated at the right of each transcript in % except for transcript 7 wich has not been reported yet. The tissue origin for each transcript was as follows : transcript 3 (t3) is the most abundant mRNA species of the cluster (96 %) is expressed in a broad variety of tissues; transcript 2 (t2) : ES cells (BB400028), adult thymus (BB033610), 13 days embryo lung (BB482014), 12 day embryo (BB092516), 8 cells embryo (BB728515), testes (AV272544), 7 days embryo (BB408501) and dendritic cell line (BY764534); transcript 1 (t1) : dendritic cells (BY206667), branchial arches (BX527136), 11 days embryo (AV123417) and adult thymus (BB033906); transcript 4 (t4) : newborn skin (BB612388), newborn whole eye (CO429499) and 16 days embryo kidney (BY060450); transcript 5 (t5) : B cells (BG145595); and transcript 6 (t6) : 16 days newborn thymus (BB197834). Transcripts represented in black (t2, t3, t7) correspond to those found in the cDNA tumor library. The amino acid sequence is indicated below each transcript. The PHD fingers are grey boxed and the nuclear/nucleolar localization signals underlined. The amino acid sequence in grey is not present in the protein and is indicated for comparison purpose.

**Figure 2.7.** Sub-cellular localization of the PHF6 protein from transcript 2 and 7. The normal PHF6 protein and the truncated form encoded by transcript 7 (PHF6 t7; see Fig. 6) were fused to the enhanced GFP (EGFP) either at the C- or N-terminal end using the pEGFP-C2 or pEGFP-N1 vector, respectively. The resulting constructs were transfected into

NIH/3T3 cells, and the cells were examined by confocal microscopy. The normal PHF6 protein and the truncated form (PHF6 t7) both in pEGFP-N1 are presented in the top and central panels, respectively. As a control, pEGFP plasmid is shown in the bottom panel.



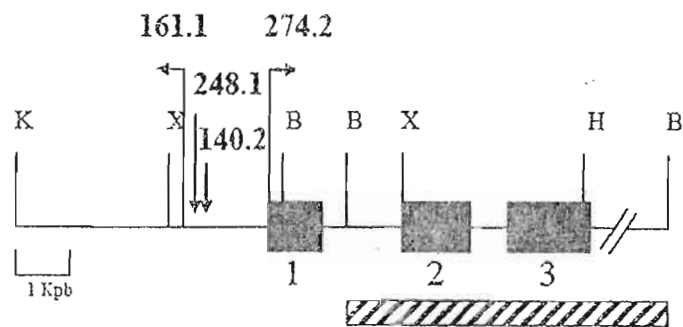
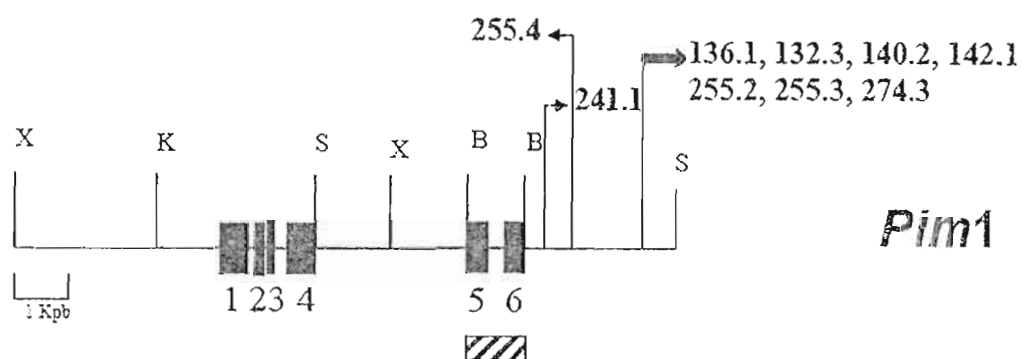
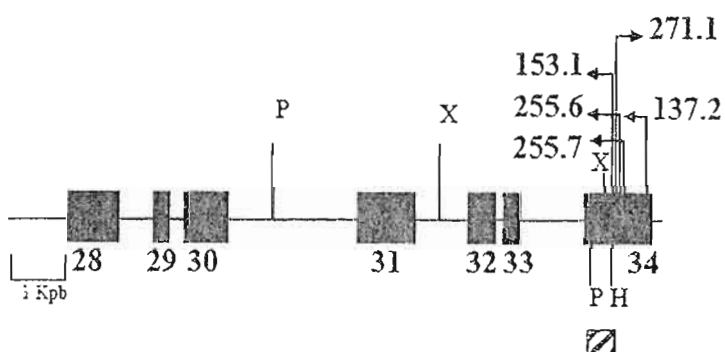
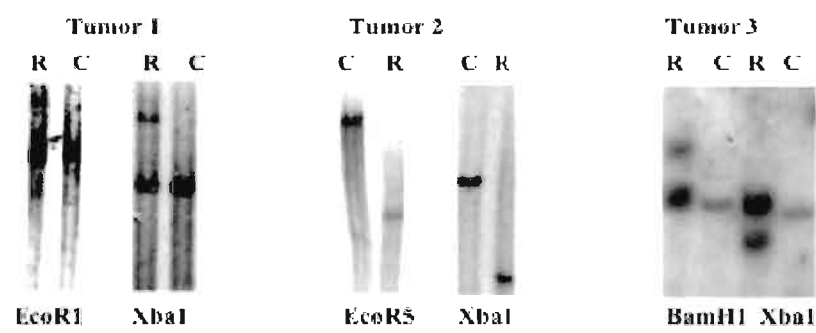
*c-myc**Pim1**Notch1*

Figure 2.1

A



B

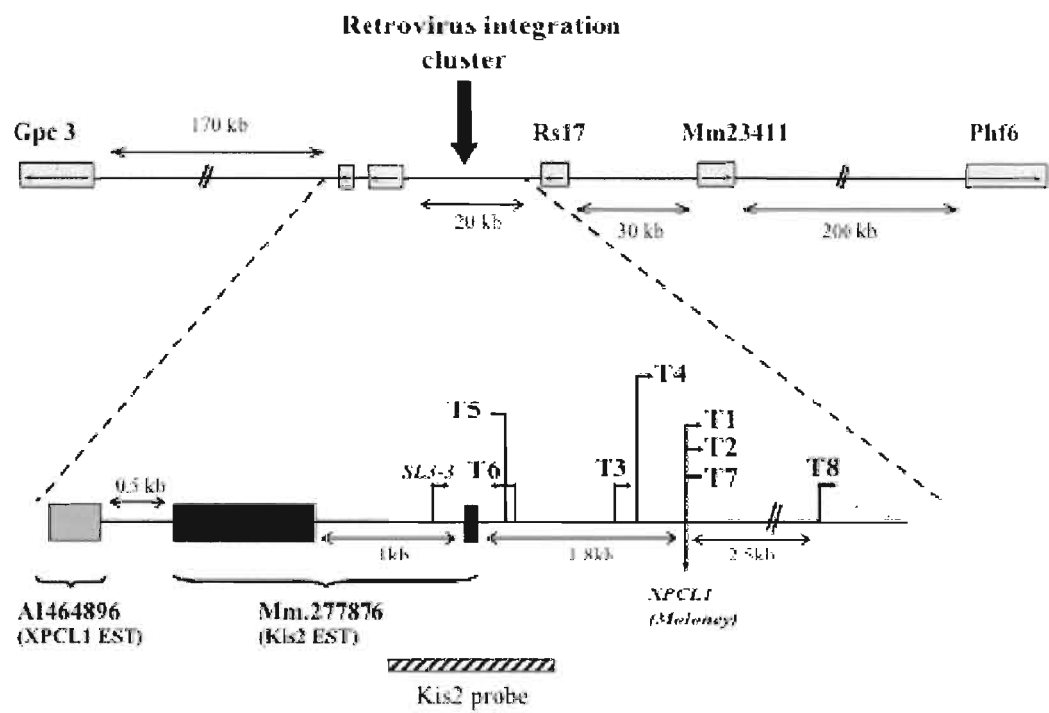
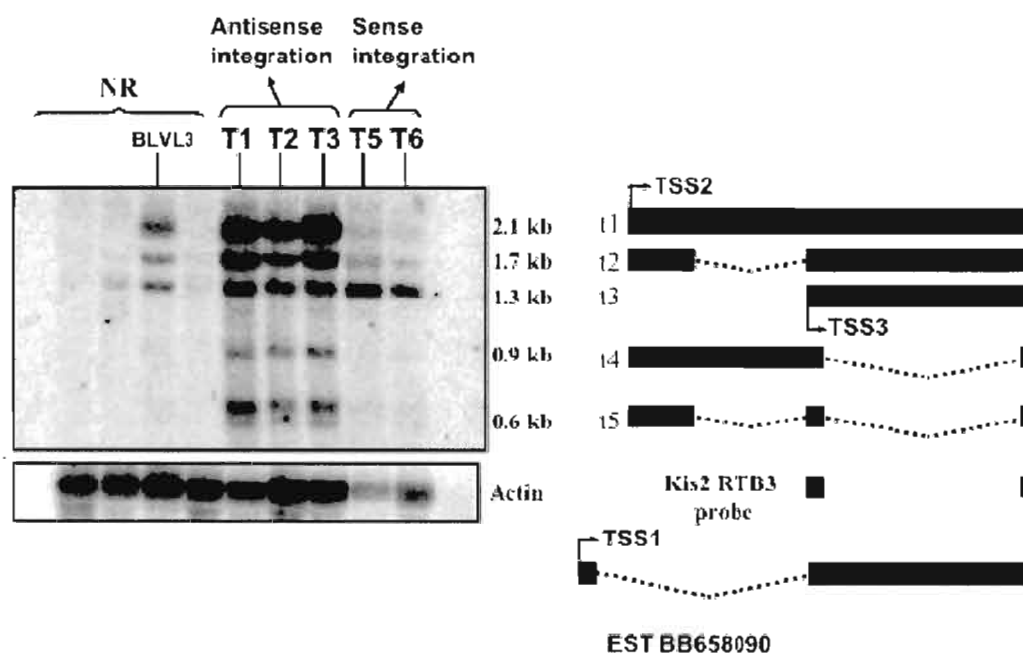


Figure 2.2

A



B

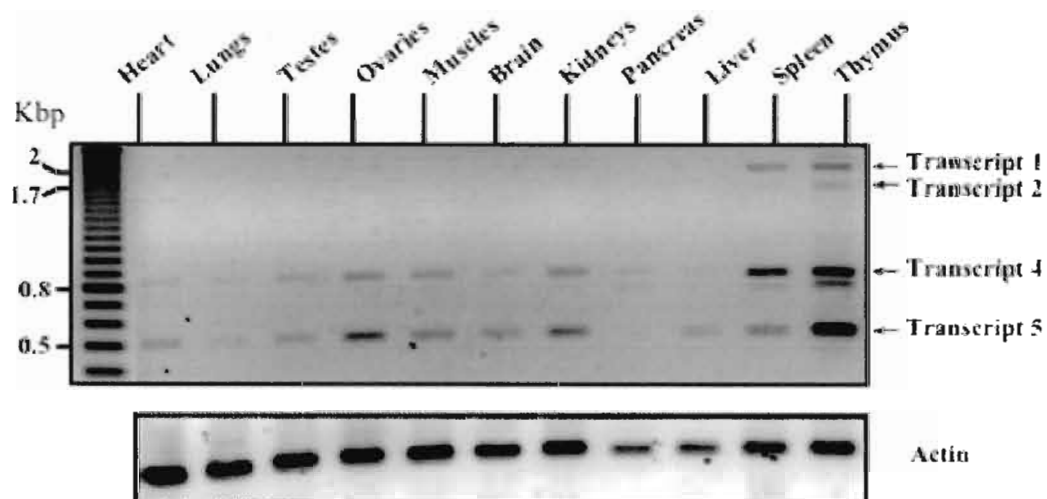
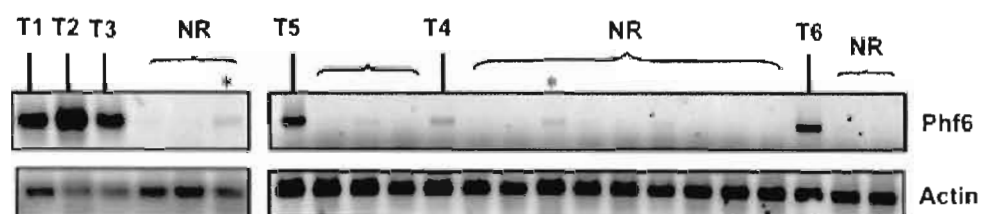
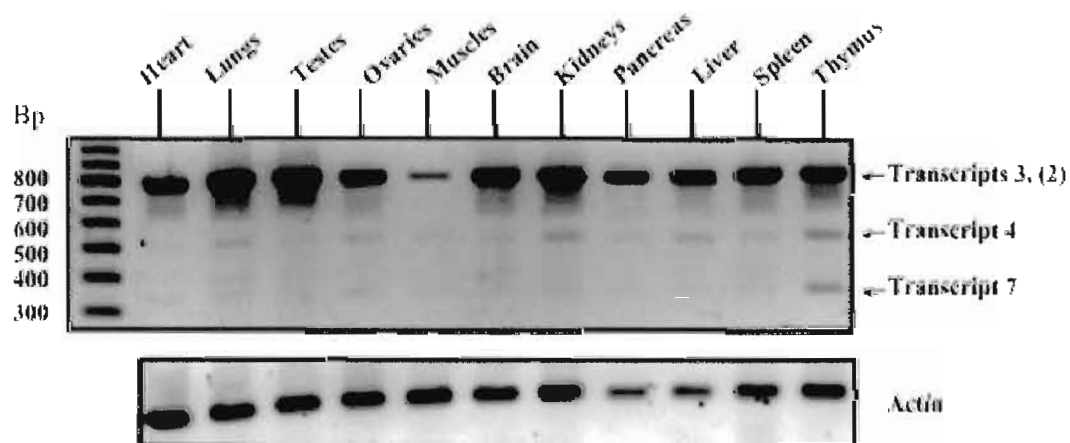


Figure 2.3



Figure 2.4

**A****B****Figure 2.5**

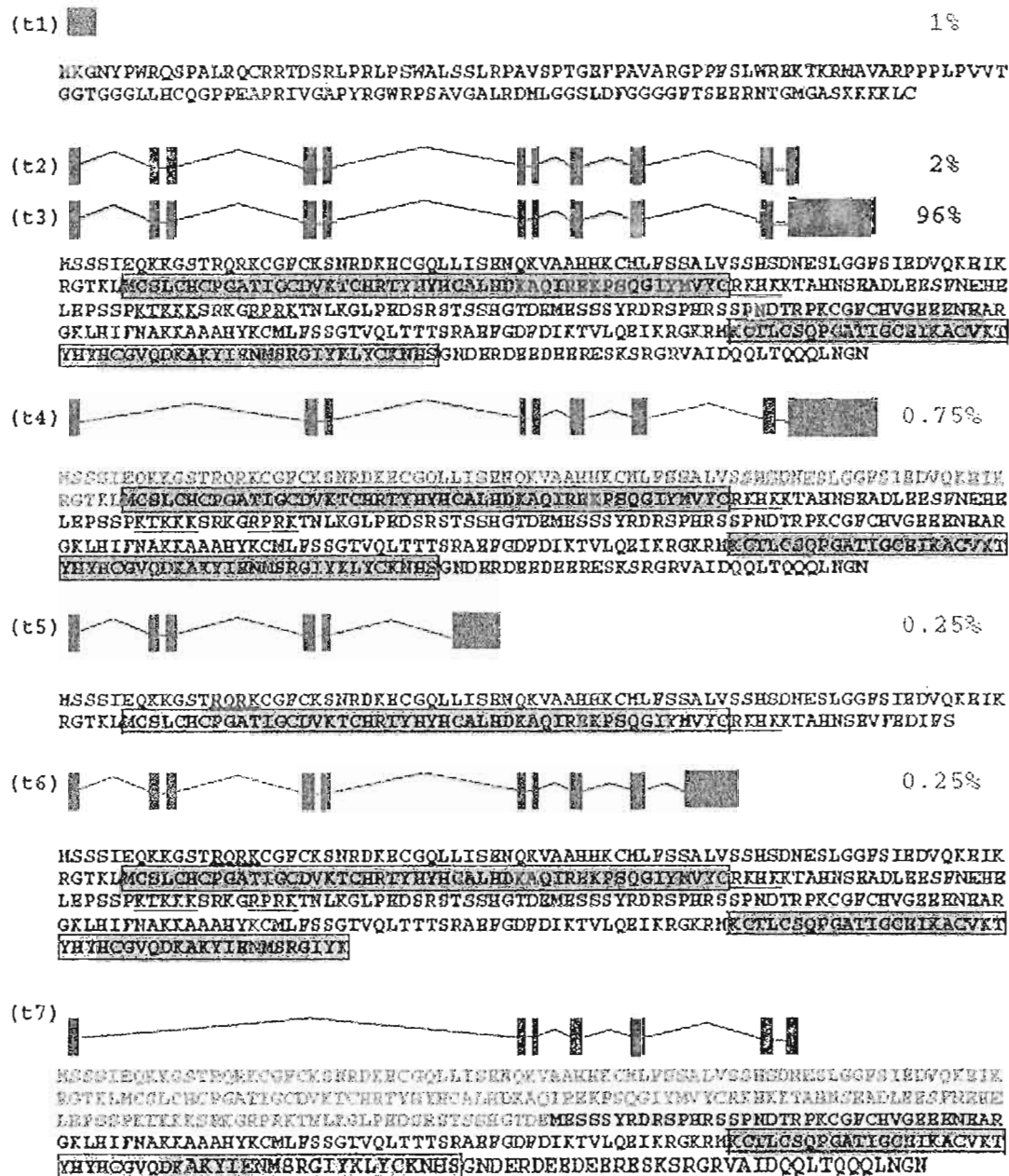
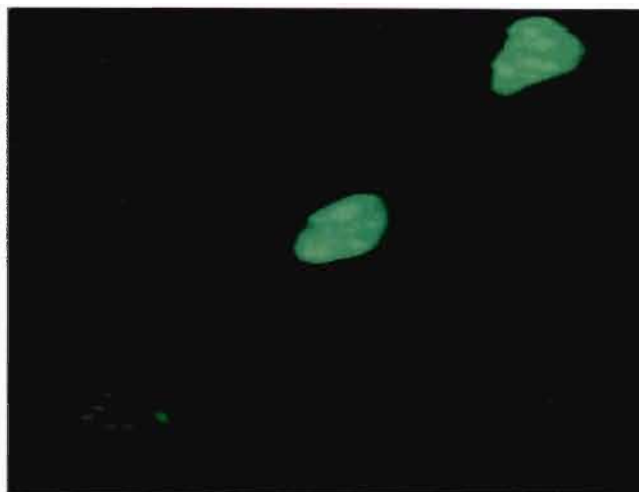
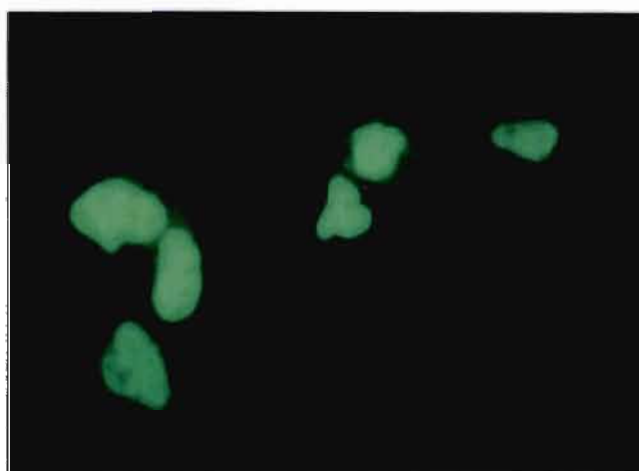


Figure 2.6

PHF6



PHF6  $\epsilon 7$



EGFP



Figure 2.7

## CHAPITRE III

### CARACTÉRISATION DES ARN NON-CODANT *KIS2* ET MISE EN ÉVIDENCE DE LEUR POTENTIEL ONCOGÈNE

#### AVANT-PROPOS

Ce deuxième chapitre concerne la caractérisation des ARN non-codant (ARNnc) *Kis2*. Lorsque j'ai commencé à travailler sur ce gène, début 2002, la notion d'ARNnc n'était pas très élaborée. En fait, bon nombre de personnes pensaient que les ARNnc n'étaient que des «artéfacts», et qu'un ARN se devait de coder pour une protéine pour avoir une fonction. Or, il existait toute une littérature rapportant l'existence d'ARNnc impliqués dans divers mécanismes tels que l'empreinte génétique (ARNnc H19, Peters *et al.* 2000) ou la régulation de la transcription (ARNnc SRA et 7SK, Lanz *et al.* 1999 et Krause *et al.* 1996). Un des ARNnc le plus étudié est XIST, impliqué dans l'inactivation du chromosome X (Brockdorff 1992). XIST fut le sujet de mon examen de synthèse, et l'occasion pour moi de découvrir une littérature fascinante tant sur l'ouverture d'esprit que sur les mécanismes invoqués. Si XIST a fait couler autant d'encre et mis à l'épreuve l'imagination des chercheurs, c'est parce qu'après 20 ans de recherches, personne ne sait vraiment comment cet ARNnc exerce sa fonction biologique. Et il en est de même pour la plupart des ARNnc connus à ce jour. Pourtant, si l'on prend en compte le contenu en intron des ARN messagers, et les données récentes issues du séquençage des génomes, 98% des ARN totaux d'une cellule seraient non-codants (Mattick *et al.* 2001, Mattick *et al.* 2003, Cawley *et al.* 2004). Il existe donc une

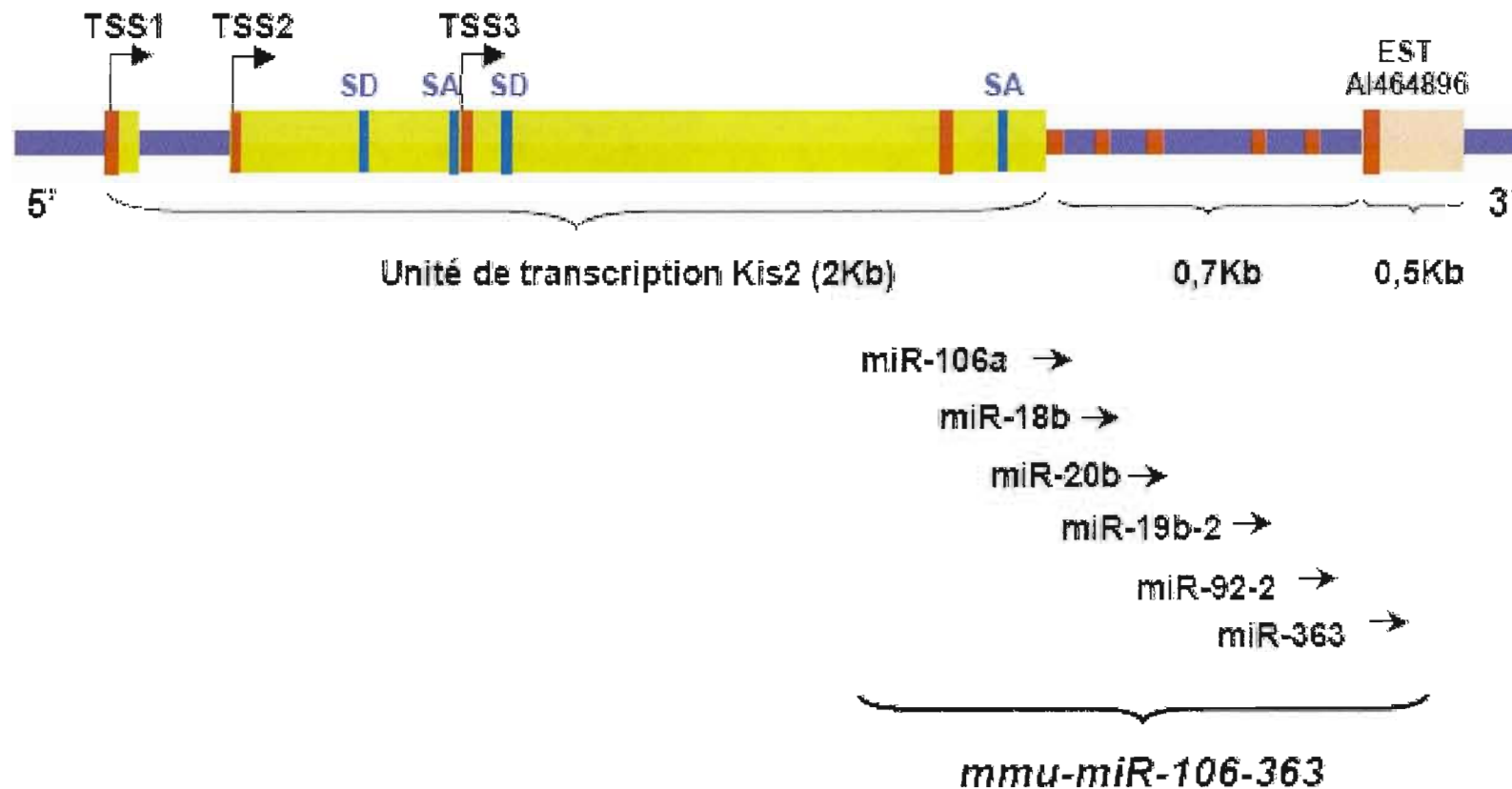


réelle dimension propre aux ARNnc, distincte des ARN messagers, qu'il reste à explorer. Avec l'émergence des microARN, le domaine des ARNnc est maintenant devenu un sujet très actuel. Et étant donné l'ampleur du phénomène microARN aujourd'hui, il paraît difficile d'imaginer que l'on soit passé à côté de ce monde pendant si longtemps. Nul doute que l'un des principaux obstacles fut la remise en question du dogme ARN/protéine.

À l'issue de ces recherches bibliographiques, j'ai décidé de poursuivre mes travaux de thèse sur la caractérisation des ARNnc *Kis2*. Je n'ai pas poursuivi le travail avec *Phf6*, car la charge de travail était trop grande pour moi seule. Par la suite, une mise à jour des banques de données devait nous rendre compte de la présence d'un cluster de microARN (miARN) en 3' du gène *Kis2* (Fig. 3.0).

Dans ce deuxième article, intitulé "Oncogenic potential of the *miR-106-363* cluster and its Implication in human T-cell leukemia", publié dans le journal "Cancer Research" (Landais *et al.* 2007), nous montrons que les ARN *Kis2* sont en fait les précurseurs du groupe de microARN *miR-106-363*. Nous montrons la surexpression de miARN de ce groupe dans les tumeurs de souris réarrangées pour *Kis2*, ainsi que dans des leucémies humaines. Nous avons de plus mis en évidence une caractéristique de leur potentiel oncogénique dans un test d'indépendance d'ancrage. Finalement, nous proposons également trois gènes cibles potentiels de *miR-106-363*, en montrant des évidences expérimentales de ce ciblage.

Au cours de ce travail, Sébastien Landry fut responsable des tests luciférase pour démontrer l'effet des miARN sur les 3'UTR des gènes cibles potentiels. Philippe Legault a réalisé le 5' RACE pour montrer que le gène *Kis2* était le pri-miARN de *miR-106-363*. Je suis responsable du schéma expérimental, de la réalisation technique des autres expériences, et de la rédaction du manuscrit.



**Figure 3.0** Situation des miARN du groupe *miR-106-363* dans la région *Kis2*. Chaque trait rouge correspond à une zone conservée entre souris, rat et humain. Chaque zone conservée en 3' du gène *Kis2* correspond à un miARN du groupe *miR-106-363*. Les traits bleus correspondent aux sites donneurs (SD) et sites accepteurs (SA) d'épissage du gène *Kis2*. La partie en jaune (et à partir de TSS2) correspond à la portion du pri-miARN de *miR-106-363* (jusqu'ici appelée ARN *Kis2*) visible en Northern-Blot.

ONCOGENIC POTENTIAL OF THE MIR-106-363 CLUSTER AND ITS  
IMPLICATION IN HUMAN T-CELL LEUKEMIA.

*SÉVERINE LANDAIS, SÉBASTIEN LANDRY, PHILIPPE LEGAULT AND  
ERIC RASSART\**

Laboratoire de biologie moléculaire, Département des sciences biologiques,  
Université du Québec à Montréal, Québec, Canada.

Running title: Oncogenic potential of the miR-106-363 cluster

**KEY WORDS:** Retroviral tagging, Oncogene activation, miRNA, leukemia,  
mylip, Rbp1-like, Hipk3.

\* Correspondance to: Eric Rassart

Département des Sciences Biologiques

Université du Québec à Montréal

Case Postale 8888 Succ. Centre-ville

Montréal, Canada . H3C-3P8

Tel. (514)987-3000 ext : 3953. Fax. (514)987-4647.

Email : [Rassart.Eric@UQAM.ca](mailto:Rassart.Eric@UQAM.ca)

### 3.1 ABSTRACT

We previously reported the identification of the *Kis2* common retrovirus integration site, located on mouse chromosome X, in RadLV-induced T-cell leukemias. Tumors with a provirus at the *Kis2* locus overexpressed a novel non-coding RNA (ncRNA) with a complex splicing pattern and no poly-A tail. Database upgrade revealed the presence of a microRNA cluster, *miR-106-363*, just downstream of the *Kis2* ncRNAs. We found that *Kis2* ncRNAs are the pri-miRNA of *miR-106-363*, and we present evidence that *Kis2* ncRNAs overexpression in mouse tumors result in miR-106a, miR-19b-2, miR-92-2 and miR-20b accumulation. We demonstrate the oncogenic potential of those miRNAs in anchorage independence assay, and confirm pri-*miR-106-363* overexpression in 46% of human T-cell leukemias tested. This overexpression contributes in rising miR-92 and miR-19 levels, as this is the case for *miR-17-92* cluster overexpression. Furthermore, we identified *Mylip*, *Rbp1*-like and possibly *Hipk3* as target genes of this miRNAs cluster, which establishes a link between these genes and T-cell leukemia for the first time.

### 3.2 INTRODUCTION

In the last few years, the RNA world has been enlightened by the emergence of microRNAs (miRNAs). Those tiny RNAs of ~21 nucleotides are involved in the attenuation of translation by base-pairing to the 3'UTR of their target mRNAs (for review see 1). These miRNAs are part of a long transcript, termed pri-miRNA, which are transcribed by the RNA polymerase II (pol II), capped and polyadenylated (2). Nearly half of the pri-miRNAs are non-coding RNAs (ncRNAs). The miRNAs are occasionally part of an intron of the ncRNA. Most of the time, ncRNAs are spliced. Once the pri-miRNA is transcribed, it is processed in the nucleus by the RNase III Drosha into smaller RNA fragments called pre-miRNA (3). These are exported in the cytoplasm where they are processed by the RNase III enzyme Dicer to produce the miRNAs (4).

In humans, it is estimated that 20 to 30% of all the genes are targeted by miRNAs, with an average of 200 genes for each miRNA (5, 6). Since a single miRNA can target several genes, and because a gene could be targeted by different miRNAs, gene expression regulation by those tiny RNA appears to be a huge complex network.

Interestingly, miRNAs appear to be critical for cellular proliferation, differentiation and apoptosis. A growing number of studies report the involvement of miRNAs in cancer and these are referred to as “oncomirs”. The first miRNAs implicated in cancer were miR-15a and miR-16-1, which are often deleted or downregulated in B-cell chronic lymphocytic leukemia (CLL) (7). Others examples came from miR-143 and miR-145, which are downregulated in colorectal cancer (8). Another well-known miRNA implicated in cancer as a tumor suppressor is *let-7*, being down-regulated in human lung cancer (9). Indeed, Johnson *et al* (10) showed that *let-7* targets the key oncogene Ras both in *C. elegans* and in human cell lines. Furthermore, the human miRNA cluster *miR-17-92* is amplified in large-B-cell lymphoma, follicular lymphoma, mantle cell lymphoma and primary cutaneous B-cell lymphoma, and acts as an oncogene collaborating with MYC (11). Interestingly, inhibition of *miR-17-92* in a cervical cancer cell line caused the upregulation of the E2F1 oncogene (12). Thus, it seems that this miRNA cluster can also act as a tumor

suppressor. Recently, miR-17-5p has been involved in breast cancer through targeting of the AIB1 oncogene (13).

In our laboratory, we use murine retroviruses as molecular bait to find oncogenes. One of our retrovirus tool is RadLV (Radiation Leukemia Virus), which specifically induces T-leukemia. A few years ago, we found a common integration site (CIS) of RadLV on mouse chromosome X called *Kis2* and we reported the overexpression of an unknown gene located nearby in the rearranged tumors (14). This gene, called *Kis2*, contains three different transcription start sites (TSS) and produces at least five major spliced RNA species, all of which are overexpressed in the analyzed tumors. All the RNA species appeared to lack an open reading frame and were consequently classified as non-coding RNAs (ncRNAs).

Recently, owing to the database update, we discovered that a cluster of six miRNAs (*miR-106a-363*) lies directly 3' to the *Kis2* ncRNAs. Interestingly, this cluster is closely related to the *miR-17-92* cluster located on mouse chromosome 14, which has been implicated in B-cell leukemia. Recently, this locus was shown to be targeted by the SL3-3 retrovirus which induces mostly T-cell leukemia in mice (15). However, despite its high homology with the *miR-17-92* cluster, the oncogenic potential of *miR-106-363* has never been studied, and it has never been implicated in leukemia.

In this article, we report that *Kis2* ncRNAs are the pri-miRNAs for the *miR-106-363* cluster in mouse. We further provide strong evidence of their oncogenic potential and their implication in human leukemias. We finally found that *myl1p*, *Rbp1-like* and possibly *Hipk3* could be considered as target genes, and consequently relates them for the first time to T-cell leukemia.

### **3.3 MATERIALS AND METHODS**

#### **3.3.1 Cell culture**

NIH3T3 cells were maintained in Dulbecco's modified Eagle's (DME) medium supplemented with 10% calf serum (Invitrogen). Medium was changed every three days, avoiding cells confluence.

#### **3.3.2 Human leukemias and normal individuals**

Human leukemia cells from blood or bone marrow were obtained from the "Banque de cellules Leucémiques du Québec" (BCLQ, Montréal, Canada). Peripheral blood mononuclear cells (PBMCs) used as control were obtained from four different adult blood samples. Adult thymus RNA was purchased from Sigma. All RNAs were extracted from cell samples with TRIzol (Invitrogen), and submitted to a DNase treatment prior to RT-PCR analysis (Ambion).

#### **3.3.3 Anchorage independence assay**

Anchorage independence assay in soft agar was performed as follow: NIH 3T3 cells were transfected with 10nM of either mmu-miR-106a, mmu-miR-19b, mmu-miR-92, mmu-miR-20b (Dharmacon) or a mix of the four miRNAs (mix assayed with two different concentration: 2,5nM or 10nM of each miRNAs) using Lipofectamin 2000 reagent (Invitrogen), according to the manufacturer's protocol. A same amount of an unraleted miRNA (Dharmacon, catalog number CN-001000-01-05) was used as a negative control. Bottom layer (0,6 % agarose in DME) was poured into 6 wells plates and 48h hours after transfections, cells were trypsinized, counted, and  $10^4$  cells were mixed with the top layer (0,3 % agarose in DME) and poured on the Bottom layer. Top layer was recovered with 1,5

mL of DME, changed twice a week. After three weeks, colonies were counted. The experiment was done twice in triplicata (n=6).

### **3.3.4 miRNAs detection by Northern-Blot**

Total RNA was extracted from frozen normal and leukemic tissues with the TRIzol reagent (Invitrogen), according to the manufacturer's protocol. For Northern analysis, 15 µg of total RNA was separated on a 12,5 % polyacrylamide gel 8M urea, and transferred to a nylon membrane (Osmonics) using a Transblot semi-dry apparatus (BioRad). Membrane was prehybridized 1 hour at 37°C in 6X SSPE (3M NaCl, 20mM EDTA, 20mM sodium phosphate pH 6.8)-5X Denhardt's-0,05M sodium phosphate pH 6.8-1mM EDTA and 0.1mg/ml denatured salmon sperm DNA. Hybridization was performed overnight at 37°C in 6X SSPE-0,05M sodium phosphate pH 6,8-5X Denhardt's-100mg/mL dextran sulfate-1mM EDTA and 0.1mg/ml denatured salmon sperm DNA. Probes used were oligonucleotides antisens to mouse miR-106a, miR-20b, miR-19b and miR-92, labeled with P<sup>32</sup> using T4 polynucleotide kinase according to the manufacturer's protocol (New England Biolabs). The membranes were washed for 10 min at 37°C in 2X SSC (1X SSC is 0.15 M NaCl-0.015 M sodium citrate), 1h at 37°C in 2X SSC-0.1% SDS, and finally for 15 min at 42°C in 0.1X SSC-0.1% SDS. Membranes were revealed by exposure and scanning with a phosphorimager (Biorad).

### **3.3.5 miRNAs detection by padlock rolling circle amplification**

Padlock rolling circle amplification was done according to Jonstrup *et al* (2006). Briefly, 0,5µg total RNA (same RNA used in RT-PCR) was mixed with 2,5 fmol of padlocks bearing a 5' phosphate specific for either hsa-miR-106a, hsa-miR-19b, hsa-miR-92 or hsa-miR-20b, heated at 65°C and allowed to cool down at room temperature. T4-DNA ligase (New England Biolabs) was added and reaction performed 2h at 37°C. Rolling circle amplification was performed upon addition of phi-29 DNA polymerase (New England



Biolabs) and 0,2mM dNTPs during 8h at 30°C followed by 10 min at 65°C. The samples were transferred to a nylon membrane (Osmonics) using a slot-blot apparatus. Membranes hybridizations were performed as described for Northern-blot. Padlocks sequences used were: pad-miR-106a (5'pTAAGCACTTTTTTATTTCCCTCAATGCTGCTGCTGTTACTACTAGTGATTTACTTGGATGTGCTACCTGCACTG3'), pad-miR-20b (5'pTGAGCACTTGTGTTATTTCCCTCAATGCTGCTGCTGTACTACTAGTGATTTACTTGGATGTCTCTACCTGCACTA-3'), pad-miR-92 (5'pC GGGACAAGTGCAATATTTATTTCCCTCAATGCTGCTGTACTACTAGTGATTTACTTGGATGTCTCAGGC-3') and pad-miR-19b (5'pTGGATTTGCACATBTTATTTCCCTCAATGCTGCTGCTGTTACTACTAGTGATTTACTTGGATGTCTTCAGTTTTGCA-3').

### 3.3.6 RT-PCR on human RNA

RT-PCR specific for *pri-miR-106-363* and *pri-miR-17-92* was performed as follow: RT (Omniscript, Qiagen) was primed on 0,5µg of total RNA with 10 pmol of 106-363 reverse primer (5'GAGGTTTCATTCACATTCAGTAACAATCTG-3'), 10 pmol of 17-92 reverse primer (5'CTAATACCACATATGAAATGTTATCAAGGC-3'), and 5 pmol of actin reverse primer (5' GACGGGGTCACCCACACTGTGCCCATCTA-3') in a 20µL reaction according to the manufacturer's protocol. PCR (Qiagen) was performed with 9µL of RT reaction for *pri-miR-106-363* and *pri-miR-17-92* amplification, and 2µL of the same RT reaction for actin detection. Forward primers used in combination with reverse primers were 106-363 forward (5'GTTGGCGACATTCCTGATCAGC-3'), 17-92 forward (5'CAGTAAAGGTAAGGAGAGCTCAATCT-3') and actin forward (5'CTAGAAGCACTTGCGGTGCACGATGGAGGG-3'). PCR program was 94°C 2 min and 35 cycles of 94°C 30 sec, 53°C 30 sec, 72°C 30 sec, followed by 10 min at 72°C. PCR products were separated on a 2 % agarose gel.

### 3.3.7 5' and 3' RACE

5' and 3' RACE were performed using First-Choice RLM-RACE kit (Ambion) according to the manufacturer's protocol. To isolate 5' ends of mouse pri-*miR-106-363* (*Kis2*), RT was primed with miR-18xba primer (5'GCTCTAGAAAGAGCCTTTGAA ACTTGTGCCGTTC-3') and nested PCR performed with oligo B primer (5'GACATCGCC AAGGCATACTCCAGGAGTGTAACC-3'). PCR program was 94°C 2 min and 35 cycles of 94°C 30 sec, 63°C 30 sec, 72°C 2 min, followed by 10 min at 72°C. Primer used for 3' RACE was Kis (5'GAGTTCTACTGCAGTG CCAGCACTTCTTAC-3'). To isolate 5' ends of human pri-*miR-106-363*, RT was primed with 106-363 reverse primer and nested PCR performed with miRxbaHum primer (5'GCTCTGAGTTACAATGACTAAATA TCAGAACT TAC-3') using the following program : 94°C 2 min and 35 cycles of 94°C 30 sec, 57°C 30 sec, 72°C 3 min, followed by 10 min at 72°C.

### 3.3.8 Luciferase assay

3'UTR's were amplified from mouse genomic DNA using high fidelity Taq polymerase (New England Biolabs), and cloned downstream of the luciferase coding sequence (XbaI site) in the pGL3 control vector (Promega). Primers used to amplify 3'UTR were: Dock4 forward (5'GTCACCTTTTCTATGTACCTGCGAT-3'), Dock4 reverse (5'CCATTTCAGCAGCAACCAACATG-3'), Rbp1-L forward (5'CAGCAGGACTTG CTAAAGC-3'), Rbp1-L reverse (5'CAGAGCCTTTGC TTACATTTGT-3'), Mylip forward (5'ATGCGTCCTGCACTCGATGGAC-3'), Mylip reverse (5'CCTACGTGACAATAA CCATAAGATCACAAAG-3'), Hipk3 forward (5'AACTCGAAAGCAGTCTATTGAG GAAGC-3'), Hipk3 reverse (5'TGTCCCTGACGGAATGTA TTAAATAAGC-3'), Mecp2 forward (5'CTTTACATAGAGCGGATTGCAAAG-3'), Mecp2 reverse (5'CATGGCC TCTAGTCTGTTGTGA-3'), Klf12 forward (5'GAAGGCTCCGGTCCAGCTAC-3'), Klf12 reverse (5'GTGCATGTGTTGCG GGTGACTG-3'). Cells were co-transfected with a mix of 0.4µg of each pGL3 construct, 0.2µg of pCMV-βGal and 20nM of each miRNA (or miRNAs

mix) using Lipofectamine 2000. The miRNA mix was assessed with either 5nM or 20nM of each miRNA. An unrelated miRNA (Dharmacon) was tested at an equal concentration (negative control). Transfected cells were lysed 48 hours post-transfection, and luciferase assay was performed as described (17), using the MLX microplate luminometer (Dynex Technologies). Luciferase activity is presented in Relative Light Units (RLU) and represents the calculated mean  $\pm$  SD of three transfected samples normalized by measuring  $\beta$ -galactosidase activity (Galacto-Light™ kit, Applied Biosystems). The experiments were done three times in triplicate.

### 3.3.9 Western-Blot

Frozen leukemic thymus or transfected cells were crushed or resuspended, respectively, in protein extraction buffer (50mM Tris-HCl [pH 7.3], 150mM NaCl, 5mM EDTA, 0.2 % Triton X-100) containing a complete protease inhibitor (Roche Applied Science). After being incubated on ice during 30 min, cells were sonicated and centrifuged 10 min at 20000g. Proteins in the supernatant were quantified using the Bio-Rad protein assay. 100 $\mu$ g of proteins were mixed to 6X loading buffer (0.35M Tris-HCl [pH 6.8], 10 % SDS, 36 % glycerol, 5 %  $\beta$ -mercaptoethanol, 0.012 bromophenol blue), and boiled 5 min to 100°C. The proteins were separated on 0,8 % SDS-polyacrylamide gel electrophoresis and electrophoretically transferred to Immobilon-P polyvinylidene difluoride membranes (Millipore). The membranes were blocked in PBS containing 0.2% Tween 20 and 5% skim milk powder for 1 h at room temperature, and incubated overnight at 4°C with antibodies diluted in PBS-Tween 5 % BSA. Mylip antibodies were kindly provided by Dr. Lindholm (Uppsala University, Biomedical Centre, Sweden) and used 1/2500, RBP1-like (BRCAA1) antibodies were purchased from Neuromics (cat. number RA19028) and used 1/1000, HIPK3 antibodies were purchased from Santa-Cruz and used 1/500. Membranes were washed in PBS-Tween and incubated with anti-rabbit/anti-mouse antibodies conjugated to peroxidase (GE healthcare, 1/20000 and 1/10000 respectively) in PBS-Tween-milk 1h at room temperature. Signal was revealed with Immobilon Western (Millipore).

### 3.4 RESULTS

#### 3.4.1 *Kis2* ncRNAs are the pri-miRNA for THE *miR-106-363* cluster

Previously, we showed that the RadLV/VL3 integration at the *Kis2* locus (mouse chromosome X, 48 989 600 -- 48 987 200 bp) gives rise to overexpression of several *Kis2* ncRNAs (14). While we were investigating the potential role of those ncRNAs in tumor development, a database upgrade allowed us to discover that six miRNAs lied directly downstream of the *Kis2* ncRNAs: miR-106a, miR-18b, miR-20b, miR-19b-2, miR-92-2 and miR-363 (fig. 3.1A). Together, they constitute the miRNA cluster, called *miR-106-363*, which is highly homologous to the *miR-17-92* cluster located on mouse chromosome 14 (fig. 3.1B). In particular, both miR-92-2 and miR-19b-2 from the *miR-106a-363* cluster are identical to miR-92-1 and miR19b-1 from the *miR-17-92* cluster. While miR-106a, miR-20b, and miR-18b share high homology with miR-17-5p, miR-20a and miR-18a from the *miR-17-92* cluster respectively, they are not identical and are thus specific of the *miR-106-363* cluster.

We were first interested to determine if *Kis2* ncRNAs could represent the pri-miRNA of this miRNA cluster. By performing 5' and 3'RACE analysis with oligonucleotides positioned in the miRNA region, we indeed found the *Kis2* TSS2 and TSS3 as 5' end (fig. 3.1C). We also found a new potential 5' end, termed TSS4, although we did not observe the corresponding transcript by Northern-blot. The 3' RACE analysis revealed that EST A1464896 was part of the 3' end of the gene (fig. 3.1A and C) and it contained the polyadenylation signal AATAAA. However, complete *miR-106-363* pri-miRNAs were never observed in Northern blots suggesting that they are rapidly and efficiently processed by Drosha. Therefore, the *Kis2* RNAs species observed on Northern blots are likely the result from Drosha processing. So far, we identified seven *Kis2* ncRNAs species, with five being overexpressed in analyzed tumors (*Kis2* t1, *Kis2* t2, *Kis2* t3, *Kis2* t4 and *Kis2* t5, fig. 3.1C) and easily detectable in Northern-Blot (14). The three others *Kis2* RNA species (t3s, t6 and t7) could be detected by RT-PCR only.

### 3.4.2 The miRNAs from the *miR-106-363* cluster are overexpressed in *Kis2*-rearranged tumors.

We next verified the expression profile of the miR-106-363 cluster in RadLV-induced tumors and found some disparities in miR-106a, miR-20b, miR-19b-2, and miR-92-2 expression. The results showed that expression levels are not equivalent for each miRNAs from the cluster. For example, miR-106a is clearly overexpressed in three tumors (139.1, 255.3 and 255.6) with a retroviral integration at the *Kis2* locus (termed rearranged tumors), while slightly overexpressed in the two other rearranged tumors (k31.5 and 132.9) (fig. 3.2A and 3.2B). In contrast, the signal for miR-20b is very weak compared to the others miRNAs, and appears only slightly overexpressed when compared to non-rearranged tumors. MiR-92 and miR-19 accumulation due to *Kis2* ncRNAs overexpression is more difficult to appreciate because these miRNAs are also products of miR-17-92. Nevertheless, miR-92 appears overexpressed in three rearranged tumors out of five (255.3, 255.6 and 139.1). For miR-19, the highest expression level is seen with tumors 255.6 and 139.1, when compared to non-rearranged tumors 159.2, k31.1, k20.2 and k24.2. The non-rearranged tumors 255.7 and 116.1 also express large levels of miR-19, but do not overexpress *Kis2* ncRNAs (data not shown). This suggests that, in this case, high levels of miR-19 could be attributable to miR-17-92. Moreover, no signals were observed for miR-18b and miR-363 (results not shown).

We conclude that the overexpression of *Kis2* ncRNAs in mice following retroviral integration could result in high accumulation of miRNAs from the miR-106-363 cluster, and in particular miR-106a. However, overexpression of both *Kis2* ncRNAs and miRNAs are not always correlated.

### 3.4.3 The *miR-106-363* cluster induces anchorage independence.

Despite its close homology with the *miR-17-92* cluster, the oncogenic potential of the *miR-106-363* cluster has never been demonstrated. To investigate the oncogenic

potential of *miR-106-363* cluster, we tested its ability to induce anchorage independence in NIH/3T3 cells. Cells transfected with either miR-106a, miR-20b, miR-19b-2 or miR-92-2 were submitted to a soft-agar assay. After three weeks, colonies could be seen with each of the tested miRNA, whereas no colonies were observed in the negative control (fig. 3.2C). Interestingly, the four tested miRNAs were not equivalent in inducing anchorage independence. We observed that miR-106a and miR-20b induced twice as much colonies than miR-92-2 or miR-19b-2 (fig. 3.2D). It should be pointed out that when all four miRNAs were transfected together, no increase in colony number was observed compared to each miRNAs transfected alone.

We conclude that, in addition to miR-106a, miR-20b, miR-19b-2 and miR-92-2 overexpression in tumors, their capacity to induce anchorage independence growth strongly supports an oncogenic potential for those miRNAs.

#### 3.4.4 Expression analysis of *miR-106-363* cluster in human T-leukemias

We next wondered if the implication of the *miR-106-363* cluster in mouse T-leukemia could be extrapolated to human T-leukemia. We looked at the expression of the human *miR-106-363* pri-miRNA in human blood or marrow samples of acute and chronic T-leukemia, and also B-cell leukemia. In addition, the expression of mature miR-106a, miR-19b-2, miR-92-2 and miR-20b were evaluated using the padlock probe (about 70 nucleotides long oligonucleotides specific to each miRNAs) and rolling circle amplification approach described by Jonstrup *et al* (16). This technique has the advantage to discriminate between highly similar miRNAs, and it could be used quantitatively. Our results show that *miR-106-363* pri-miRNA is clearly amplified compared to controls in six tumors out of thirteen (LALT3b, LALT1j, LALT, LALT3j, LAPTj1, LPBb2, and LMr) (fig. 3.3A and B). When we looked at the expression of individual miRNAs, we found that miR-92-2 and miR-19b-2 were detected in all the tested tumors. These results are consistent with the fact that miR-92 and miR-19b are also products of the *miR-17-92* cluster. When looking at the *miR-17-92* pri-miRNA expression, it appears more regularly expressed among both tumors and controls than

the *miR-106-363* cluster. Thus, *miR-17-92* should be responsible for the miR-92 and miR-19b signals in all tumors where *miR-106-363* is not detected. However, the signals of both miR-92-2 and miR-19b-2 appear stronger in those tumors that show *miR-106-363* overexpression (LALT, LALT3b, LALT3j, or LAPTj1, for example). This suggests that *miR-106-363* overexpression contributes to the increase of miR-92 and miR-19b as well. No signal was detected for miR-20b. Surprisingly, miR-106a was detected in only two tumors out of six overexpressing *miR-106-363*. We expected an increased expression of miR-106a in those six tumors, as it was the strongest signal seen by Northern blot in mouse tumors (Figure 3.2A). This result highlights the existence of a regulation for mature miRNAs production as already mentioned by others (see discussion).

We conclude that overexpression of *miR-106-363* very likely participates in the oncogenic process leading to human leukemias (as previously demonstrated for *miR-17-92*). However, overexpression of the pri-miRNA of a miRNAs cluster is not synonymous with overexpression of all the miRNAs produced from the cluster and careful examination of each tumor is required.

### 3.4.5 Identification of potential target for the *miR-106-363* cluster

To find targets related to the *miR-106-363* cluster in a tumoral context, we hypothesize that those miRNAs should exert their oncogenic potential by targeting tumor suppressor mRNAs. Since four miRNAs from the *miR-106-363* cluster are expressed at the same time in our tumors, we focussed on targets common to miR-106a, miR-19b-2, miR-20b, and miR-92-2. We thought that each of them could target the same genes in order to cooperate for an efficient inhibition. We first adopted an *in silico* approach and worked with the prediction websites from the computational biology center of the Memorial Sloan-Kettering Cancer Center ([www.microrna.org](http://www.microrna.org)) and Targetscan ([www.targetscan.org](http://www.targetscan.org)). Targets predictions are based upon the miRNA “seed”, which corresponds to nucleotide 2 to 8 starting from the miRNA 5’ end. Indeed, high quality pairing of the miRNA seed seems to be important for stability of the miRISC complex (19, 20) as well as for biological function (21, 22). Based on this property, one should consider that any target prediction related to miR-

106a should be envisaged for miR-20b as well, as they both possess the same seed (fig 3.1B and fig. 3.4). Among the common target predictions (146 genes predicted), we selected genes based on their known tumor suppressor activity, or genes whose downregulation could lead to tumor development (Table 3.1). Furthermore, we limited our choice in this study to the highest prediction scores, and target sites perfectly conserved between human, mouse and rat. We next tested the ability of all four miRNAs to target the 3'UTR of the selected gene in a luciferase assay. We found that three out of six tested genes showed significantly reduced luciferase activity: *Mylip* (Myosin regulatory Light chain Interacting Protein), *Hipk3* (Homeodomain Interacting Protein Kinase 3) and *Rbp1-like* (Retinoblastoma-Binding Protein 1-like)(fig. 3.5A). The latter was selected because of a high prediction score even though this transcript was not targeted in its 3'UTR by miR-92-2. The three other genes (*Mecp2*, *Klf12* and *Dock4*) did not seem to be targeted by miR-106a, miR-20b, miR-19b-2 or miR92-2 in our assay (results not shown). In contrast to our predictions, not all four tested miRNAs could modulate gene expression (fig. 3.4). In the case of *Rbp1-like*, only miR-106a/miR-20b exerted a significant inhibitory effect. This could be correlated with the presence of two targets sites for miR-106a/miR-20b in its 3'UTR (Fig 3.4). Surprisingly, miR-19b-2 appeared to have no effect on *Rbp1-like* 3'UTR despite the prediction of a target site (fig 3.4). As expected, no targeting by miR-92-2 was observed. All four miRNAs displayed an inhibitory effect on *Mylip* 3'UTR (Fig 3.5A). Similarly, miR-92-2 efficiently targeted *Hipk3* 3'UTR, while miR-106a/miR-20b and miR-19b-2 seemed to have a lower but significant effect.

As already observed with the anchorage independence growth assay, no synergistic effect from miRNA targeting was observed on *Mylip* and *Hipk3* 3'UTR when all four miRNAs were present. Similar results were obtained with a more concentrated mix (results not shown).

### **3.4.6 Endogenous MYLIP and RBP1-like are downregulated in NIH/3T3 by *miR-106-363* cluster.**

To confirm our results from luciferase assays, we verified if endogenous *Mylip*, *Rbp1-like* and *Hipk3* protein levels could be affected in NIH/3T3 by the miRNAs from the



*miR-106-363* cluster. Transfections of miR-106a and miR-20b clearly downregulated levels of MYLIP and RBP1-like proteins (fig. 3.5B). However, miR-92-2 and miR-19b-2 appeared to be less efficient in inhibiting endogenous MYLIP production, while they seemed to have no effect on RBP1-like production. This lack of effect of miR-92-2 and miR-19b-2 on Rbp1-like expression correlates with the results observed in our luciferase assay. Unfortunately, the HIPK3 protein could not be detected with commercial antibodies and thus, miRNA inhibition could not be confirmed at the protein level. However, the correlation between the luciferase results and Mylip and Rbp1-like protein levels strongly suggests that Hipk3 is likely to be a real target of *miR-106-363*.

According to our results, we conclude that *Mylip* and *Rbp1-like* are true targets of the *miR-106-363* cluster. More specifically, *Rbp1-like* appeared only targeted by miR-106a/miR-20b whereas *Mylip* could be targeted by the four miRNAs with miR-106a/miR-20b being more efficient.

### **3.4.7 MYLIP and RBP1-like are downregulated in RadLV tumors overexpressing *miR-106-363* cluster.**

We next decided to verify the level of MYLIP and RBP1-like proteins in our RadLV-induced tumors that overexpressed the *miR-106-363 cluster*. Western blot analysis for MYLIP and RBP1-like showed lower protein levels in those tumors that overexpressed the four miRNAs compared to negative tumors (fig. 3.5C). As miR-106a is strongly overexpressed in the tumors, these results are consistent with our previous results which show a more efficient inhibition of MYLIP and RBP1-like with miR-106a/20b (Fig 3.5B). Interestingly, tumor 139.1 which showed the highest miRNAs levels also exhibited the lowest protein level for both MYLIP and RBP1-like (fig 3.2A).

Targeting of Rb by miR-106a in human has already been reported (23). Since the miR-106a target site in Rb 3'UTR is perfectly conserved between human and mouse, we also looked at the expression of RB. We observed varying levels of protein among the tumors analyzed (fig. 3.5C). Low levels of RB were detected not only in those tumors overexpressing miR-106a but also in tumors which did not. Thus, if the RB protein can be

targeted by miR-106a as demonstrated by others, our results show that miR106a overexpression is not always correlated with RB reduction. As RB is often downregulated in tumors, miRNAs are probably not the only mechanism explaining a reduction in RB protein level.

In summary, our results strongly suggest that *Myliip* and *Rbp1-like* RNAs are targeted by the *miR-106-363* cluster in mouse tumors associated with its overexpression.

### 3.5 DISCUSSION

In this article, we report that the overexpressed Kis2 ncRNAs resulting from the retroviral integration within the locus are the pri-miRNAs for the *miR-106-363* cluster in mouse, thus revealing this cluster for the first time as an oncogene in T-leukemia. Indeed, we demonstrated its ability to induce anchorage independence in NIH/3T3 cells, and its overexpression in 46% of human T-cell leukemias analyzed.

Our results are also supported by reports indicating that miRNAs from the *miR-106-363* cluster are overexpressed in some solid tumors, as for example miR-106a and miR-92-2 in colon and prostate (23). Furthermore, miR-106a, miR-92 and miR-19 have recently been reported to be overexpressed in K562 and U937 leukemia cell lines (24). All together, these results reveal the oncogenic potential of the *miR-106-363* cluster both in leukemias but also in others cancers types.

One of the most striking clues concerning the oncogenic abilities of the *miR-106-363* cluster is its close homology with the *miR-17-92* cluster, and in particular miR-19 and miR-92 which are 100% identical in both clusters. Our results show that while *miR-17-92* seems to be constitutively expressed in normal cell, this is not the case for *miR-106-363* which could be strikingly overexpressed in tumors, thus contributing in rising up miR-19 and miR-92 levels. As a consequence, when miR-92 or miR-19 are detected, one should not only consider the *miR-17-92* implication but *miR-106-363* as well. This contribution of two distinctively regulated loci in producing identical miRNAs has also been described for the miR-15/miR-16 cistron (25). Similar to our observations with *miR-17-92* and *miR-106-363*, one locus expresses constitutively *miR-15-16*, whereas the other is induced during development only.

When miRNAs are not identical, they could share high homology, as miR-17-5p and miR-106a, or miR-106a and miR-20b for example. This redundancy between several miRNAs is not yet understood, and has consequences on target predictions. To date, most predictions are based upon 5' miRNAs seed, which tends to be a common signature between several miRNAs, as for example miR-17-5p, miR-106a, and miR-20b. Consequently, target predictions are nearly identical for those three miRNAs, and one could speculate that overexpression of the *miR-106-363* cluster would have nearly the same effects than *miR-17-*

92 overexpression. However, if highly similar miRNAs can share similar targets, it is likely that their small nucleotidic differences could confer some specificity to each miRNAs in terms of targeting. Some studies showed clearly the importance of miRNAs 5'end, but also highlight modulatory effects from the 3'end (21, 22).

Furthermore, some interesting questions arise from miRNAs clusterization. Do those miRNAs collaborate in regulating the same gene? Alternatively, do they regulate different genes in identical or distinct pathways? Since cancer is a multi-step process involving collaboration of several genes and pathways, miRNAs clusters appear to be ideal oncogenic candidates. To date, most of the miRNAs implicated in cancer are indeed part of miRNAs cluster, as miR-15/miR-16 (7), miR-372/miR-373 (26) or miR-143/miR-145 (8). We show in our study that miR106a/miR-20b, miR-92 and miR-19 could commonly target distinct sites of the Mylip gene. However, in contrast to *let-7* targeting of Ras where RAS inhibition is proportional to the number of *let-7* targets sites (10), this "common targeting" does not seem to occur synergistically in our hands when the four miRNAs are presents in the cell. Common targeting could operate at other levels, and each miRNA quantity could also be important.

In light of this, microarray expression analysis of miRNAs often revealed differential expression among miRNAs from clusters. Our results also revealed quantity differences between miRNAs produced from the *miR-106-363* cluster, with miR-106a being not detected in a subset of human leukemias despite the presence of the pri-miRNA. In the same way, miR-363 and miR-18 are not detected in mice analyzed tumors, and miRNAs from the *miR-106-363* cluster are not overexpressed at the same level. This highlights the emergent regulation hypothesis of miRNA clusters processing. This is demonstrated in the recent study by Thomson *et al* (25) who show that the processing of pri-miRNAs by Drosha is a key regulation step in miRNAs production. The block could occur not only at the Drosha step, but each miRNA from a same cluster could be differentially processed.

As a miRNA could target a lot of genes at the same time, and as the overexpression of a miRNA cluster gives rise to accumulation of several miRNAs, it is difficult to identify important targets in tumorigenesis. A combination of targeting events by miRNAs cluster is probably the more realistic view, and protein arrays technology the better way to handle it. Nevertheless, we attempted to identify *miR-106-363* targets among existing predictions that

could be related to tumor development, and we found that Mylip, Rbp1-like and most likely Hipk3 are targets genes of this cluster.

MYLIP is an ERM-like protein that interacts with myosin regulatory light chain and its only known function is to inhibit neurite outgrowth (27). However, ESTs of this gene are found in a broad variety of tissues, including thymus and lymph nodes. Its ERM (Ezrin-Radixin-Moesin) domain retained our attention, because of its close homology with the ERM domain of the tumor suppressor NF2/Merlin. ERM proteins such as Merlin link the actin cytoskeleton to membrane bound proteins, and are involved in regulation of cell motility, cell shape, contact inhibition and cell attachment (28). We show here for the first time that Mylip represents a novel interesting candidate gene whose function as a tumor suppressor in human leukemias deserve to be investigated.

The Rbp1-like gene is already linked to cancer, although this information is still controversial. First, RBP1-like is recognized by IgG antibodies isolated from a breast cancer patient and used as a marker (29). Second, it shares high homology with RBP1 (Retinoblastoma associated protein 1), and both proteins are part of the mSin3A corepressor complex involved in histone deacetylation (30). Moreover, overexpression of either RBP1 or RBP1-like caused a profound inhibition of cell proliferation and induced expression of a senescence marker (31), thus conferring tumor suppressor activities to RBP1-like. Our results concerning Rbp1-like targeting by miR-106a/20b in tumors provides an additional clue to its potential tumor suppressor role. Furthermore, Rbp1-like targeting by miR-17-5p/20a should be considered in tumors overexpressing this miRNA.

Finally, HIPK3 has been identified as a Fas interacting kinase (32). Overexpressing HIPK3 in mammalian cells impaired Fas ligand-induced activation of Jun NH2-terminal kinase, without affecting cell death. However, Möller *et al* (33) reported that Hipk3 associates in the nucleus with its highly homologous tumor suppressor Hipk2 to form the Hipk domains. We believe that it would be interesting to investigate the tumoral suppressor aspect of Hipk3, and further confirm its targeting by the *miR-106-363* cluster.

In conclusion, our work highlights the importance of *miR-106-363* cluster in leukemia, and raises important questions about its close relationship to *miR-17-92* cluster.

Also, it provides three new candidate genes with tumor suppressor potential, which should be considered as targets of the *miR-17-92* cluster as well.

## ACKNOWLEDGEMENT

We would like to thank Dr Benoît Barbeau for its helpful discussion. We also thank Dr. Lindholm (Uppsala University, Biomedical Centre, Sweden) for the generous gift of Mylip antibodies. All the human leukemia samples were kindly provided by the “Banque de cellules Leucémiques du Québec” (BCLQ, Montréal, Canada). This work was supported by grant MOP-37994 from the Canadian Institutes of Health Research.

### 3.6 REFERENCES

1. **Ouellet DL, Perron MP, Gobeil LA, Plante P, Provost P.** MicroRNAs in Gene Regulation: When the Smallest Governs It All. *J. Biomed. Biotechnol* 2006, 4: ID69616.
2. **Cai X, Hagedorn C.H, and Cullen B.R.** Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* 2004, 10:1957–1966.
3. **Lee Y, Ahn C, Han J, et al.** The nuclear RNase III Drosha initiates microRNA processing. *Nature* 2003, 425:415-419.
4. **Bernstein E, Caudy A.A, Hammond S.M and Hannon G.J.** Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 2001, 409:363-366.
5. **Lewis B.P, Burge C.B, and Bartel D.P.** Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005, 120:15-20.
6. **Krek A, Grun D, Poy M.N et al.** Combinatorial microRNA target predictions. *Nat. Genet* 2005, 37:495-500.
7. **Calin G.A, Liu C.G, Sevignani C et al.** MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc Natl Acad Sci U S A* 2004, 101(32): 11755–11760.
8. **Michael M.Z, O'Connor S.M, van Holst Pellekaan N.G, Young G.P & James R.J.** Reveals that *mir-143* and *mir-145* RNA levels are often reduced in colorectal cancer. *Mol. Cancer Res* 2003, 1:882–891.
9. **Takamizawa J, Konishi H, Yanagisawa K et al.** Reduced Expression of the *let-7* MicroRNAs in Human Lung Cancers in Association with Shortened Postoperative Survival. *Cancer Res* 2004, 64:3753–3756.
10. **Johnson S.M, Grosshans H, Shingara J et al.** *RAS* Is Regulated by the *let-7* MicroRNA Family. *Cell* 2005, 120:635-647.
11. **He L, Thomson J.M, Hemann M.T et al.** A microRNA polycistron as a potential human oncogene. *Nature* 2005, 435:828-833.

12. **O'Donnell K.A, Wentzel E.A, Zeller K.I, Dang C.V & Mendell J.T.** c-Myc-regulated microRNAs modulate E2F expression. *Nature* 2005, 435 :839-843.
13. **Hossain A, Kuo M.T , and Saunders G.F.** Mir-17-5p Regulates Breast Cancer Cell Proliferation by Inhibiting Translation of *AIB1* mRNA. *Mol. Cell. Biol* 2006, 26:8191-8201.
14. **Landais S, Quantin R, and Rassart E.** RadLV common integration at the *Kis2* locus: Simultaneous overexpression of a novel non-coding RNA and of the proximal *Phf6* gene. *J. Virol* 2005, 79:11443-11456.
15. **Wang C.L, Wang B.B, Bartha G et al.** Activation of an oncogenic microRNA cistron by provirus integration. *Proc Natl Acad Sc* 2006, 103 (49):18680–18684.
16. **Jonstrup S.P, Koch J, and Kjems J.** A microRNA detection system based on padlock probes and rolling circle amplification. *RNA* 2006, 12(9):1747-1752.
17. **Cavanagh M.H, Landry S, Audet B et al.** HTLV-I antisense transcripts initiating in the 3'LTR are alternatively spliced and polyadenylated. *Retrovirology* 2006, 3: 15.
18. **Mulder L.C.F, Chakrabarti L.A, and Muesing M.A.** Interaction of HIV-1 Integrase with DNA Repair Protein hRad18. *J. Biol. Chem* 2002, 277: 27489-27493.
19. **Khvorova A, Reynolds A & Jayasena S.D.** Functional siRNAs and miRNAs exhibit strand bias. *Cell* 2003, 115:209–216.
20. **Schwarz D.S, Hutvagner G, Du T, Xu Z, Aronin N and Zamore P.D.** Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 2003, 115:199–208.
21. **Brennecke J, Stark A, Russell R.B, and Cohen S.M.** Principles of MicroRNA–Target Recognition. *PLOS Biology* 2005, 3:404-418.
22. **Doench J.G & Sharp P.A.** Specificity of microRNA target selection in translational repression. *Genes Dev* 2004, 18:504–511.
23. **Volinia S, Calin G.A, Liu C.G et al.** A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sc* 2006, 103: 2257–2261.
24. **Yu J, Wang F, Yang G.H et al.** Human microRNA clusters: Genomic organization and expression profile in leukemia cell lines. *Biochem Biophys Res Commun* 2006, 349(1):59-68.



25. **Thomson J.M, Newman M, Parker J.S, Morin-Kensicki E.M, Wright T and Hammond S.M.** Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes & Dev* 2006, 20: 2202-2207.
26. **Voorhoeve P.M, Le Sage C, Schrier M *et al.*** A Genetic Screen Implicates miRNA-372 and miRNA-373 As Oncogenes in Testicular Germ Cell Tumors. *Cell* 2006, 124:1169–1181.
27. **Olsson P.A, Korhonen L, Mercer E.A, and Lindholm D.** MIR is a novel ERM-like protein that interacts with myosin regulatory light chain and inhibits neurite outgrowth. *J. Biol. Chem* 1999, 274(51) :36288-36292.
28. **Bretscher A, Reczek D and Berryman M.** Ezrin: a protein requiring conformational activation to link microfilaments to the plasma membrane in the assembly of cell surface structures. *Journal of Cell Science* 1997, 110: 3011-3018.
29. **Cui D, Jin G, Gao T *et al.*** Characterization of BRCA1 and its novel antigen epitope identification. *Cancer Epidemiol Biomarkers Prev* 2004 13(7): 1136-1145.
30. **Andersen J.S, Lyon C.E, Fox A.H *et al.*** Identification and characterization of three new components of the mSin3A corepressor complex . *Mol. Cell. Biol* 2003, 23 (10): 3456-3467.
31. **Binda O, Roy J.S, and Branton P.E.** RBP1 family proteins exhibit SUMOylation-dependent transcriptional repression and induce cell growth inhibition reminiscent of senescence. *Mol. Cell. Biol* 2006, 26(5): 1917-1931.
32. **Rochat-Steiner V, Becker K, Micheau O, Schneider P, Burns K, and Tschopp J.** FIST/HIPK3: a Fas/FADD-interacting Serine/Threonine Kinase that Induces FADD Phosphorylation and Inhibits Fas-mediated Jun NH 2 -terminal Kinase Activation. *J Exp Med* 2000, 192(8):1165-1174.
33. **Möller A, Sirma H, Hofmann T.G *et al.*** PML Is Required for Homeodomain-interacting Protein Kinase 2 (HIPK2)-mediated p53 Phosphorylation and Cell Cycle Arrest but Is Dispensable for the Formation of HIPK Domains. *Cancer Res* 2003, 63: 4310-4314.

### 3.7 FIGURE LEGENDS

**Figure 3.1:** (A) Genomic map of the *Kis2* locus on mouse chromosome X. The RadLV integration cluster is represented by the vertical arrow. Genes and EST clusters of the region are schematized by grey boxes. The previously identified *Kis2* gene is represented by dark boxes. The arrow in the box indicates the transcription orientation. The six miRNAs of the *miR-106-363* cluster in 3' of *Kis2* ncRNA are represented by arrows (miR-106a, miR-18b, miR-20b, miR-19b-2, miR-92-2, miR-363). (B) miRNAs organization of the mouse miR-17-92 and *miR-106-363* clusters, and sequence comparison between miRNAs from each cluster. Boxes correspond to each miRNA stem-loop with grey part corresponding to miRNA position in the stem-loop. (C) *Kis2* ncRNAs species identified as pri-miRNA for *miR-106-363* cluster. Four Transcription Start Sites (TSS1 to 4) have been identified and are represented by angled-arrows. The grey part represents the pri-miRNAs region containing the six miRNAs which are represented by arrows (miR-106a to miR-363). Black boxes correspond to the pri-miRNA part which could be revealed in Northern blot, while the dark gray box at the extreme 3' end corresponds to EST AI464896 (Landais et al 2005). Shaded *Kis2* transcripts were only detectable by RT-PCR.

**Figure 3.2:** (A) Expression analysis of miRNAs from the *miR-106-363* cluster in mice RadLV induced tumors. Northern blot showing overexpression of miR-106a, miR-20b, miR-19b-2 and miR-92-2 in *Kis2*-rearranged tumors compared to non-rearranged tumors. The 5.8S RNA serves as the loading control. (B) Quantification of signal relative to Northern-Blot in (A). (C) Transfection of each of miR-106a, miR-19b-2, miR-20b and miR-92-2 (10nM) induces anchorage independence in a soft agar assay (bottom panel, miR-106a). Top panel shows the negative control (Cells transfected with an unrelated miRNA). (D) Colony counting from the soft agar assay (experiment done twice in triplicate). Mix corresponds to a mix of the four miRNAs in equal amount each at 2,5nM.

**Figure 3.3:** (A) Expression analysis of the *miR-106-363* cluster in human T-leukemias. Pri-miRNAs from *miR-106-363* and *miR-17-92* have been amplified by RT-PCR, actin serves as a loading control (top 3 panels). miR-106a, miR-19, miR-92 were detected by a padlock

rolling circle amplification assay. U6 serves as a loading control (bottom 4 panels). RNA from thymus and PBMCs serves as normal control. Tumor legend is: LPT (blood): prolymphocytic T leukemia; LPB (b1 from blood and b2 from marrow): pro-lymphocytic B leukemia; LM (blood): mantle leukemia; LLC (blood): chronic lymphocytic B leukemia; LALT (1j from marrow and others from blood): acute lymphocytic T leukemia; LAPT (blood): acute pre-lymphocytic T leukemia; LAL-L1 (marrow): acute lymphocytic T leukemia. (B) Quantification of signals observed in (A).

**Figure 3.4:** Schematization of *Hipk3*, *Myliip*, and *Rbp1-like* 3'UTR. The miR-106a, miR-19b-2 and miR-92-2 target sites are represented by gray boxes. Sequence complementarity between target sites (gray) and the miRNAs (dark) is shown.

**Figure 3.5:** (A) Analysis of different 3'UTR in a luciferase assay. Targeting of the 3'UTR of *Myliip*, *Rbp1-like* and *Hipk3* by miR-106a, miR-20b, miR-19b-2 and miR-92-2 was tested in a luciferase assay in NIH/3T3 cells. As negative control, the experiment was done with an unrelated miRNA. Each miRNA has been assessed individually, or together (2.5nM mix). Experiment were done twice in triplicata (B) Western Blot of the RBP1-like and MYLIP protein levels in NIH/3T3 cells transfected with either miR-106a, miR-19, miR-92, miR-20b, a mix of the four miRNAs each at 2.5nM (mix) or with an unrelated miRNA (neg ctrl). Experiment were done in triplicata (C) Western blot of MYLIP, RBP1-like and RB protein levels in tumors overexpressing (\*) or not the *miR-106-363* cluster. GAPDH serves as a loading control.

**Table 3.1 : *miR-106-363* target genes predicted and tested in our luciferase assay.**

<b>Gene name</b>	<b>Function</b>
<b>Dock4*</b>	GTPase activator
<b>Myli1</b> (Myosin regulatory Light chain Interacting Protein)	Identified as a neurone outgrowth inhibitor
<b>Rbp1-like*</b> (Retinoblastoma Binding Protein 1 like)	Inhibits cell proliferation
<b>Hipk3</b> (Homeodomain Interacting Protein Kinase 3)	Fas/FADD-interacting serine/threonine kinase
<b>Klf12</b> (Kruppel-like factor 12)	Transcriptional repressor
<b>Mecp2</b> (Methyl CpG binding Protein 2)	Methylation mediated silencing

\* Gene known as tumor suppressor.

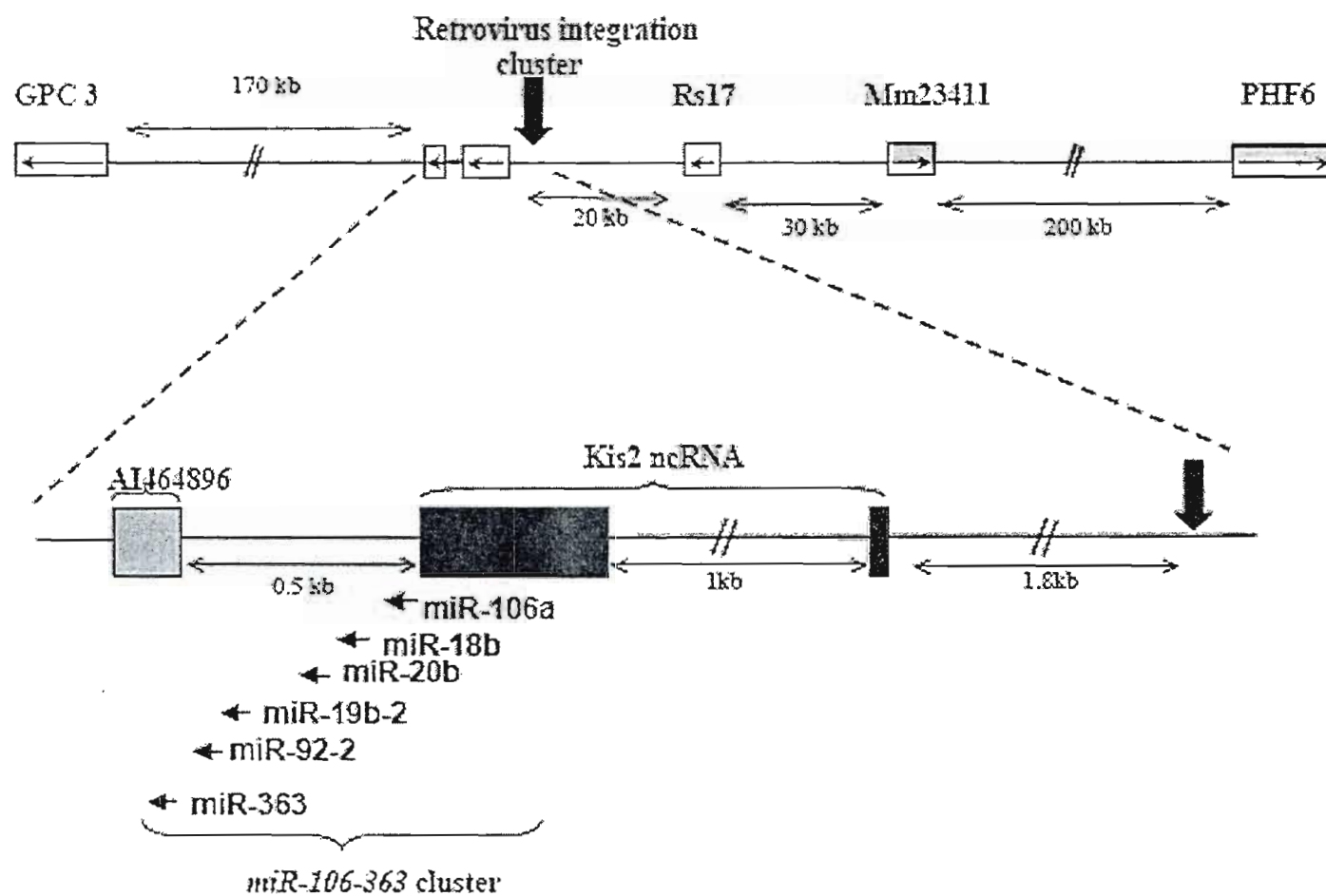


Figure 3.1A

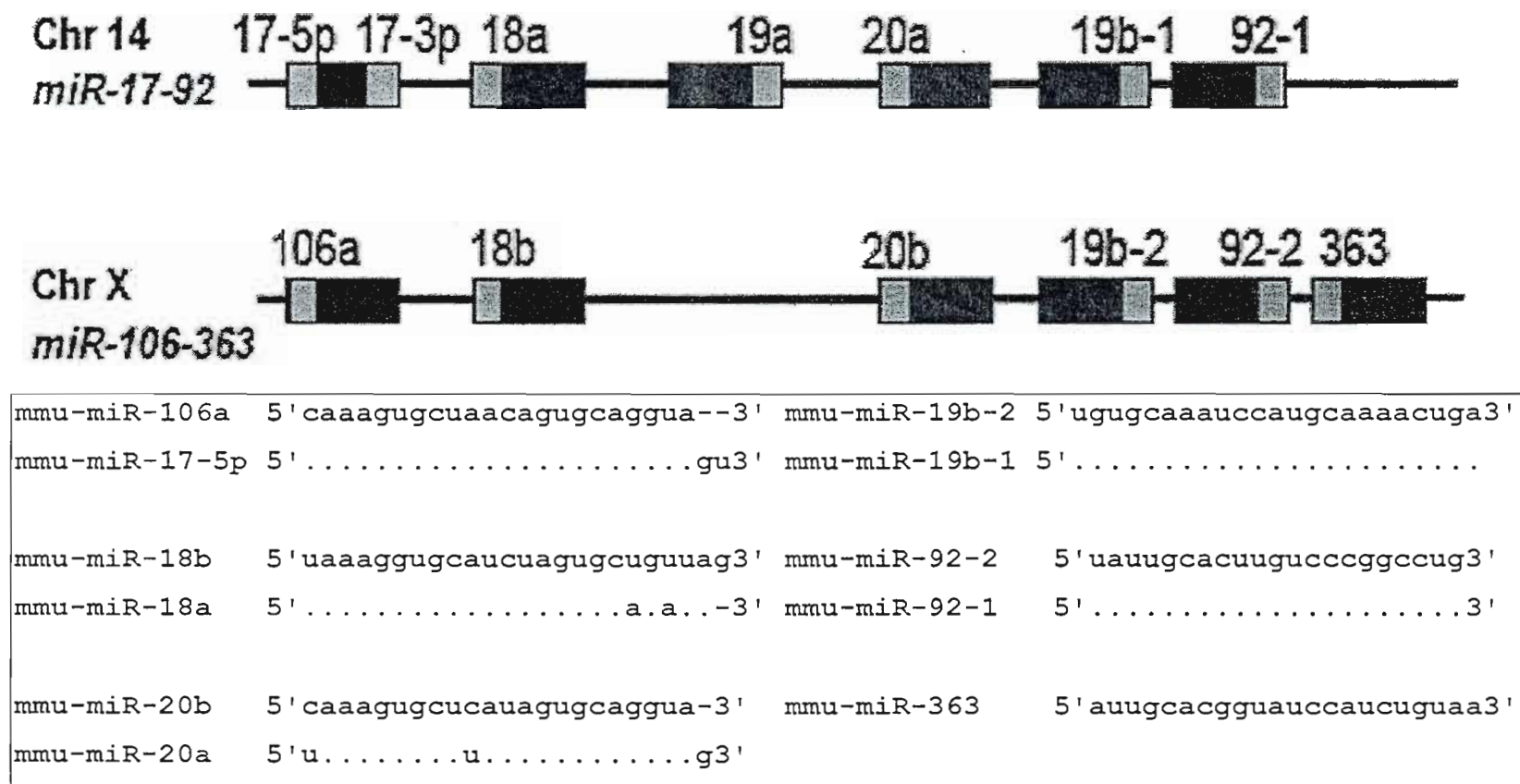


Figure 3.1B

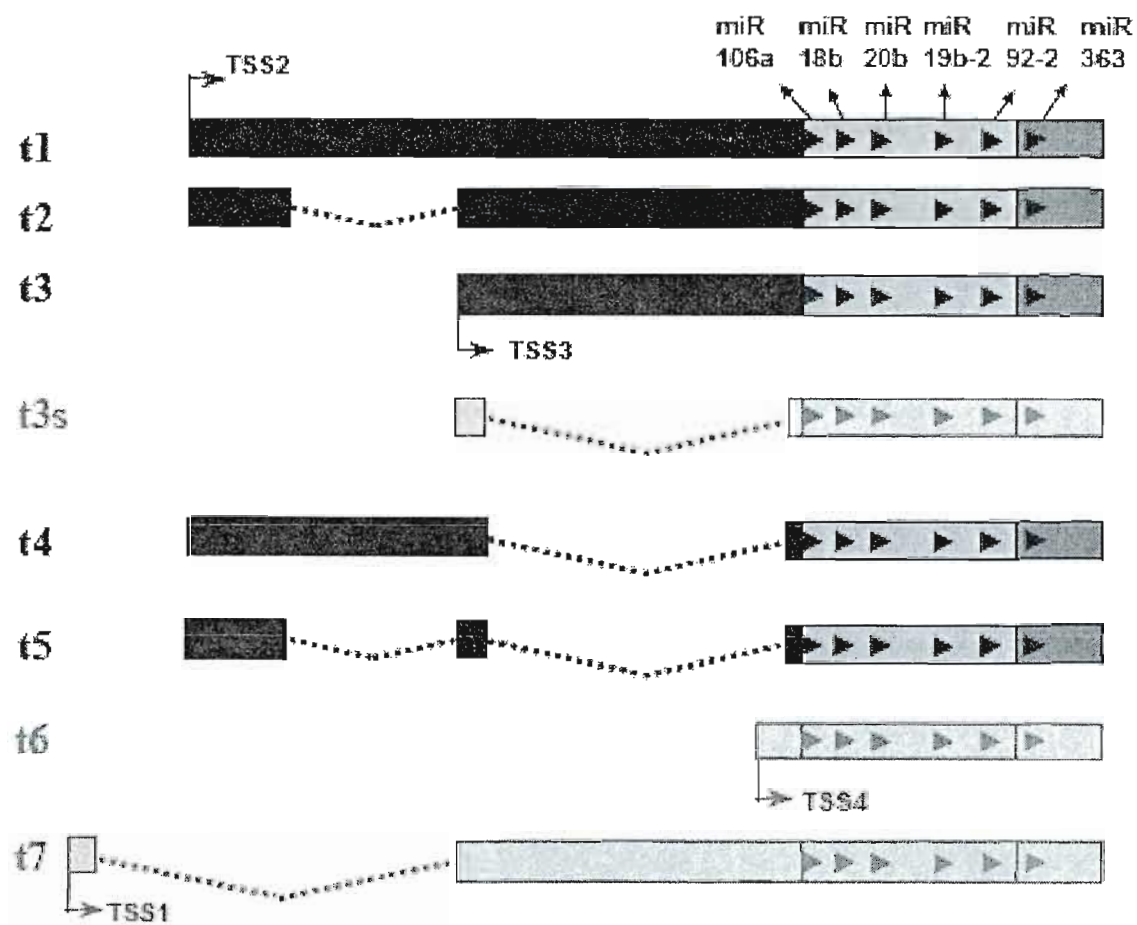


Figure 3.1C

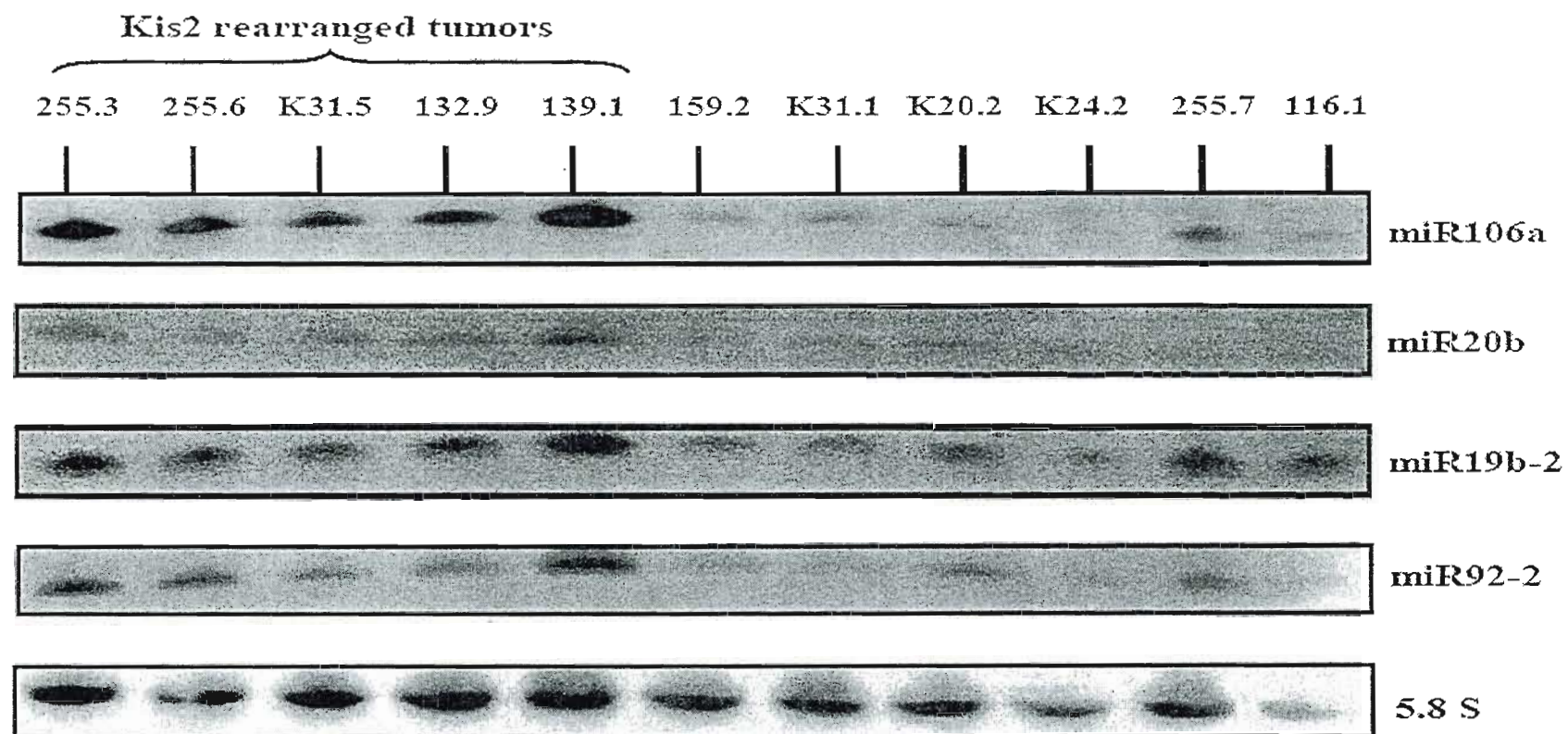


Figure 3.2A



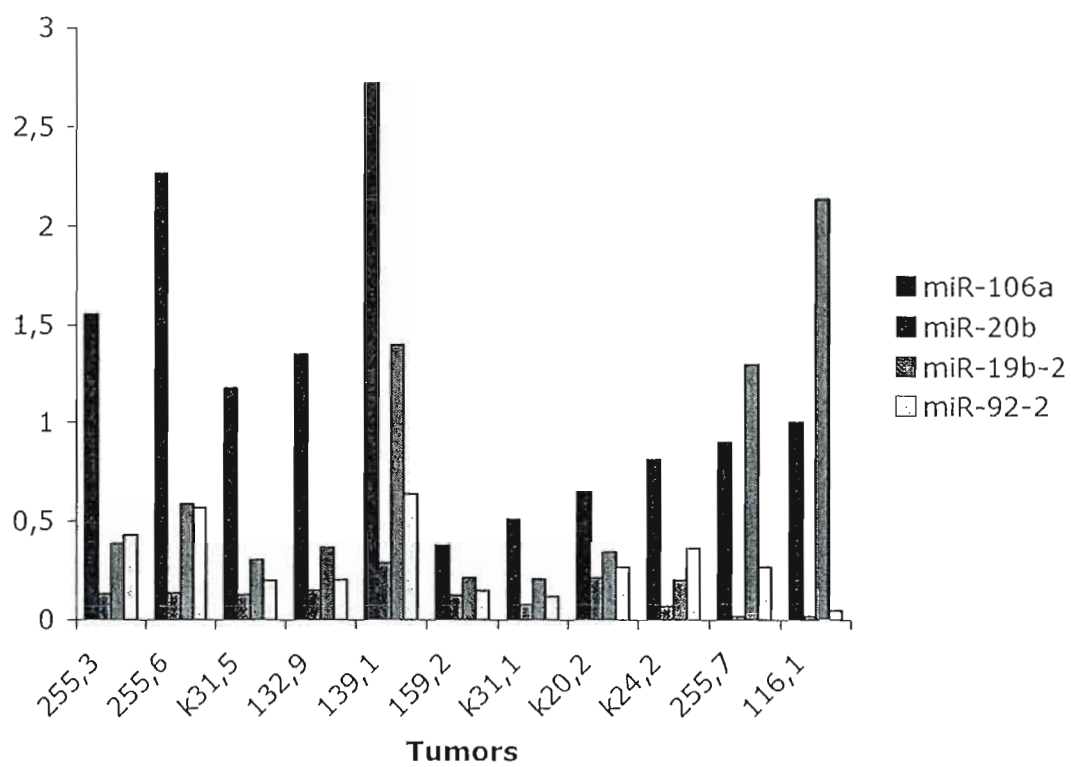


Figure 3.2B

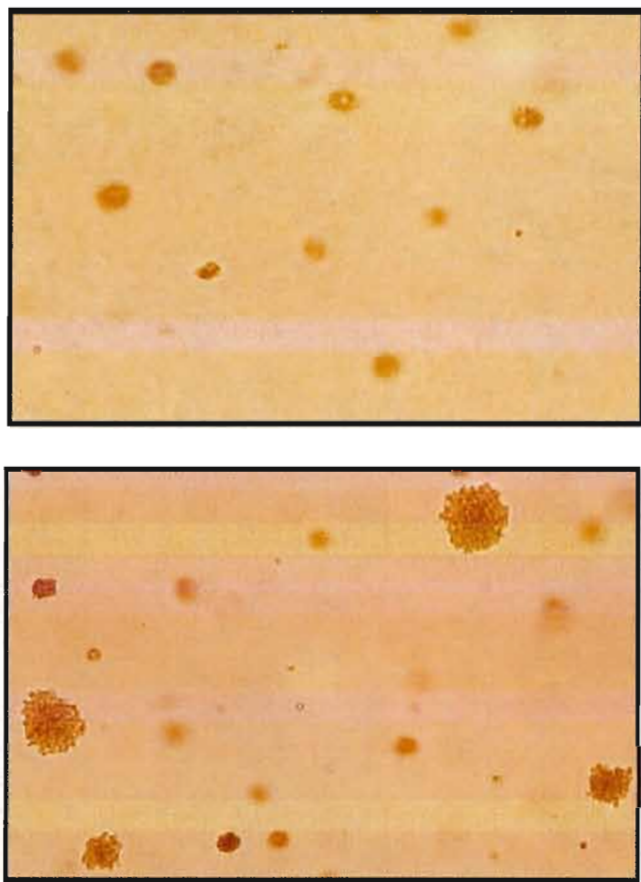


Figure 3.2C

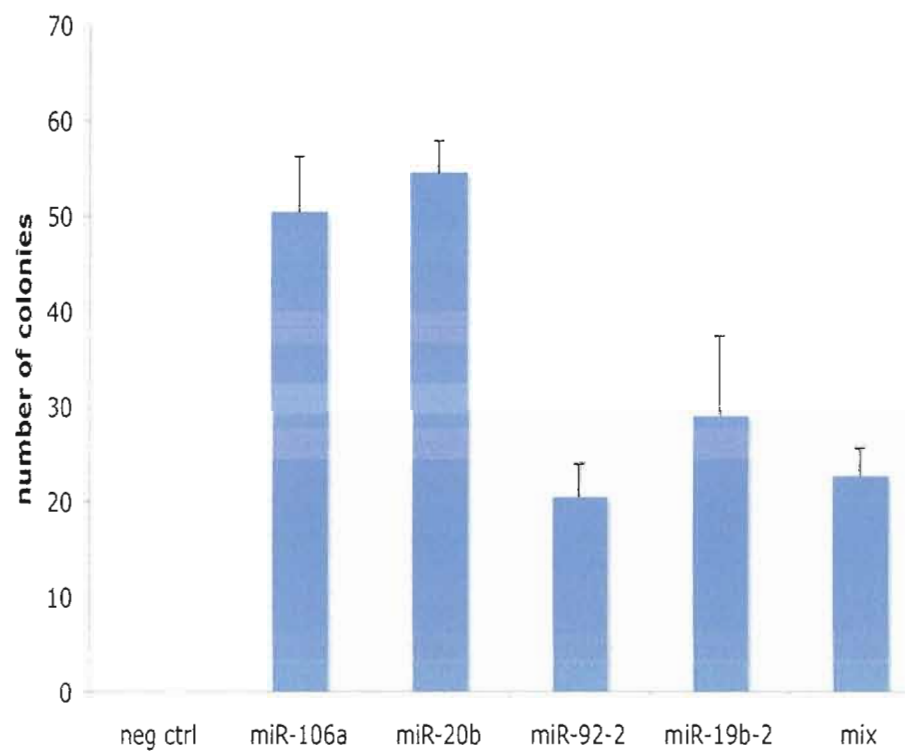


Figure 3.2D

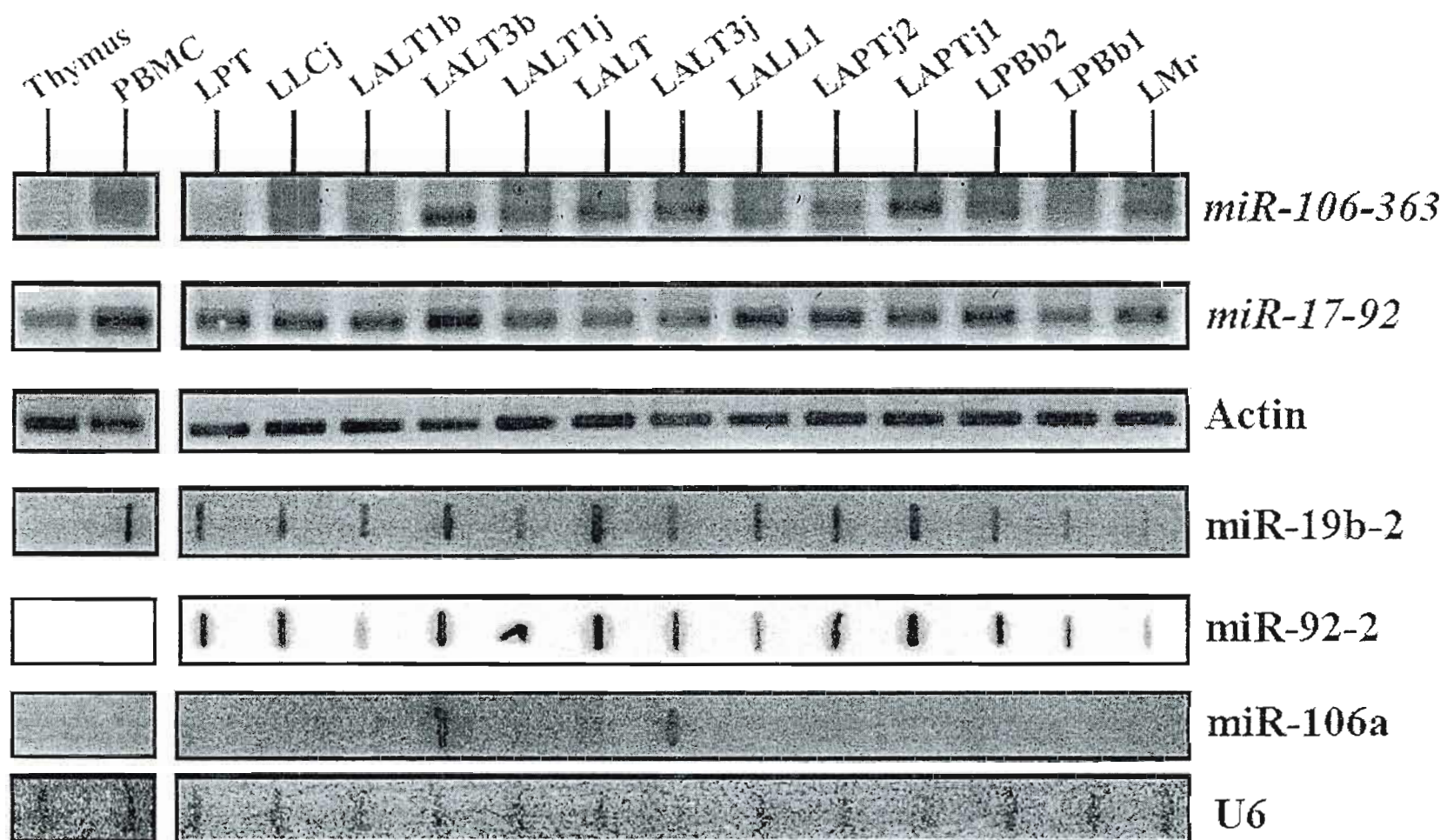


Figure 3.34

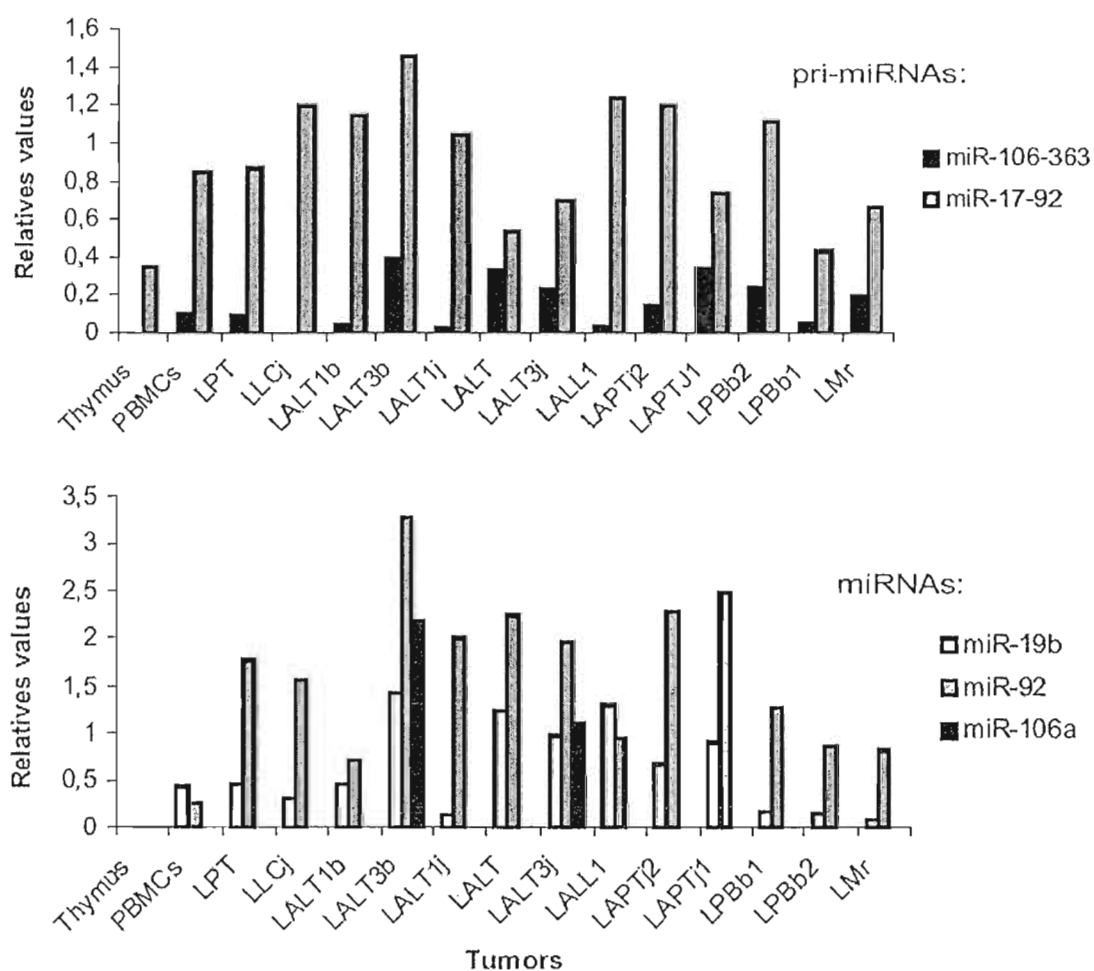


Figure 3.3B

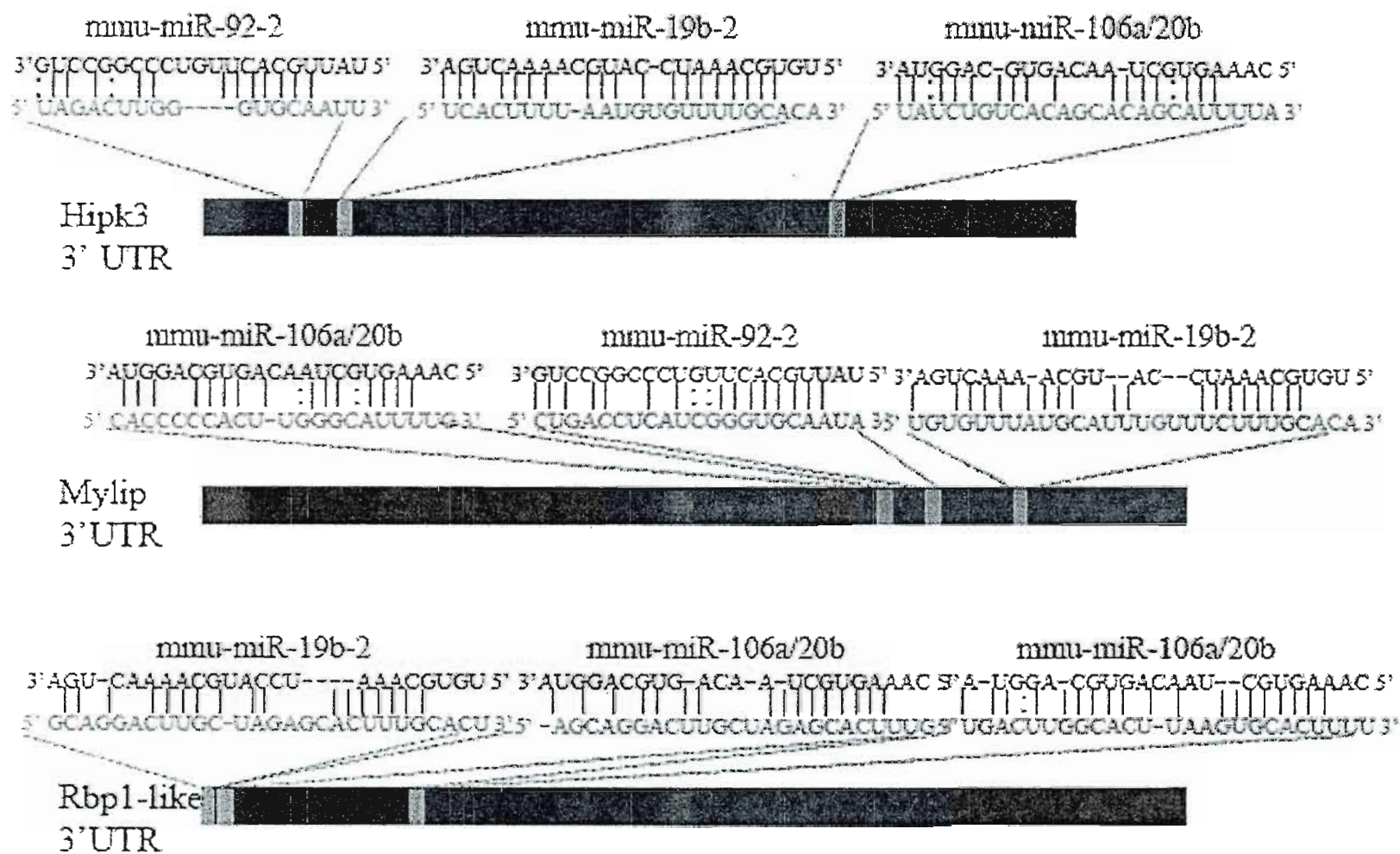


Figure 3.4

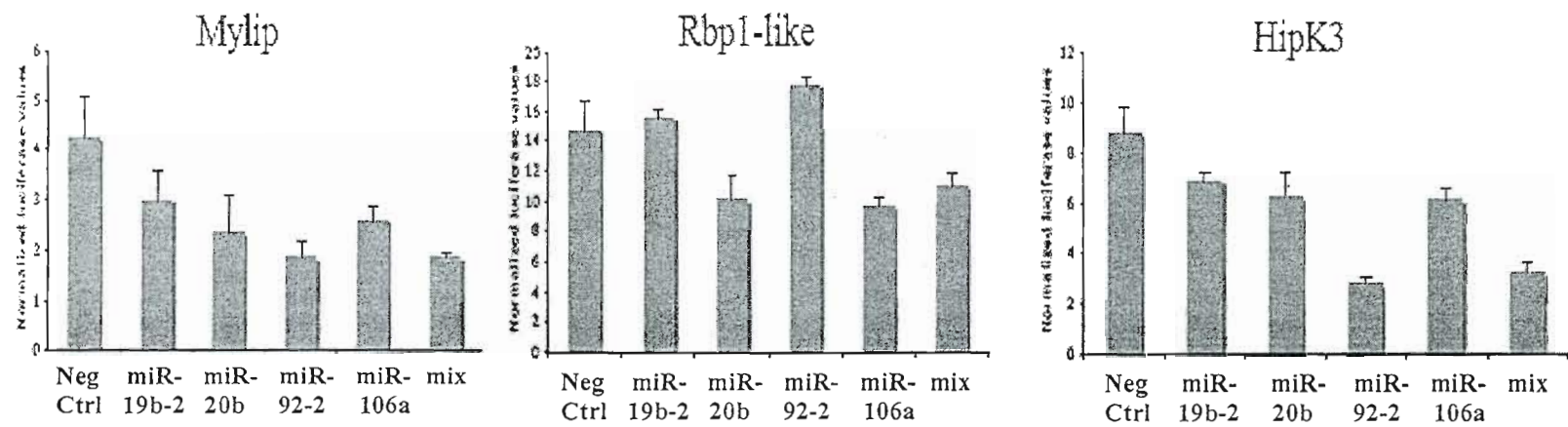


Figure 3.5A

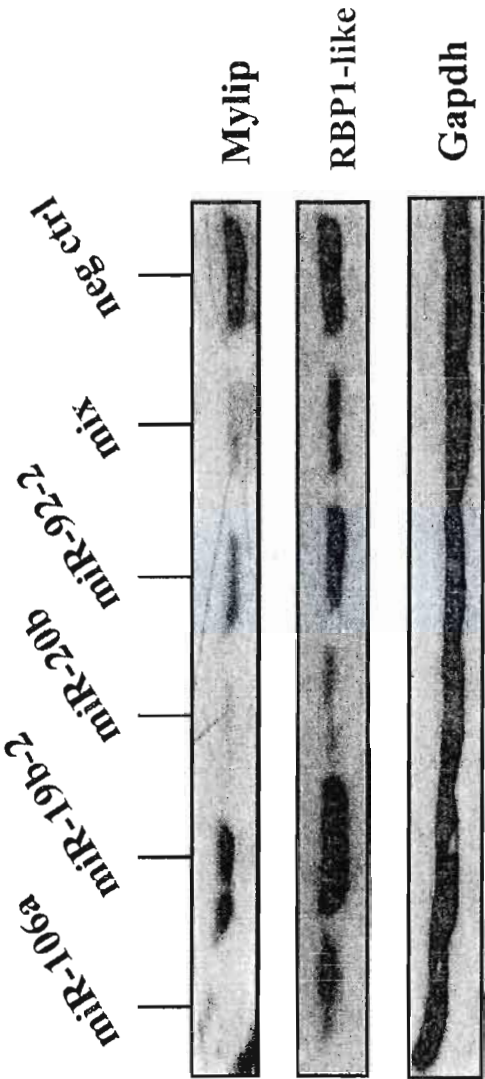


Figure 3.5B



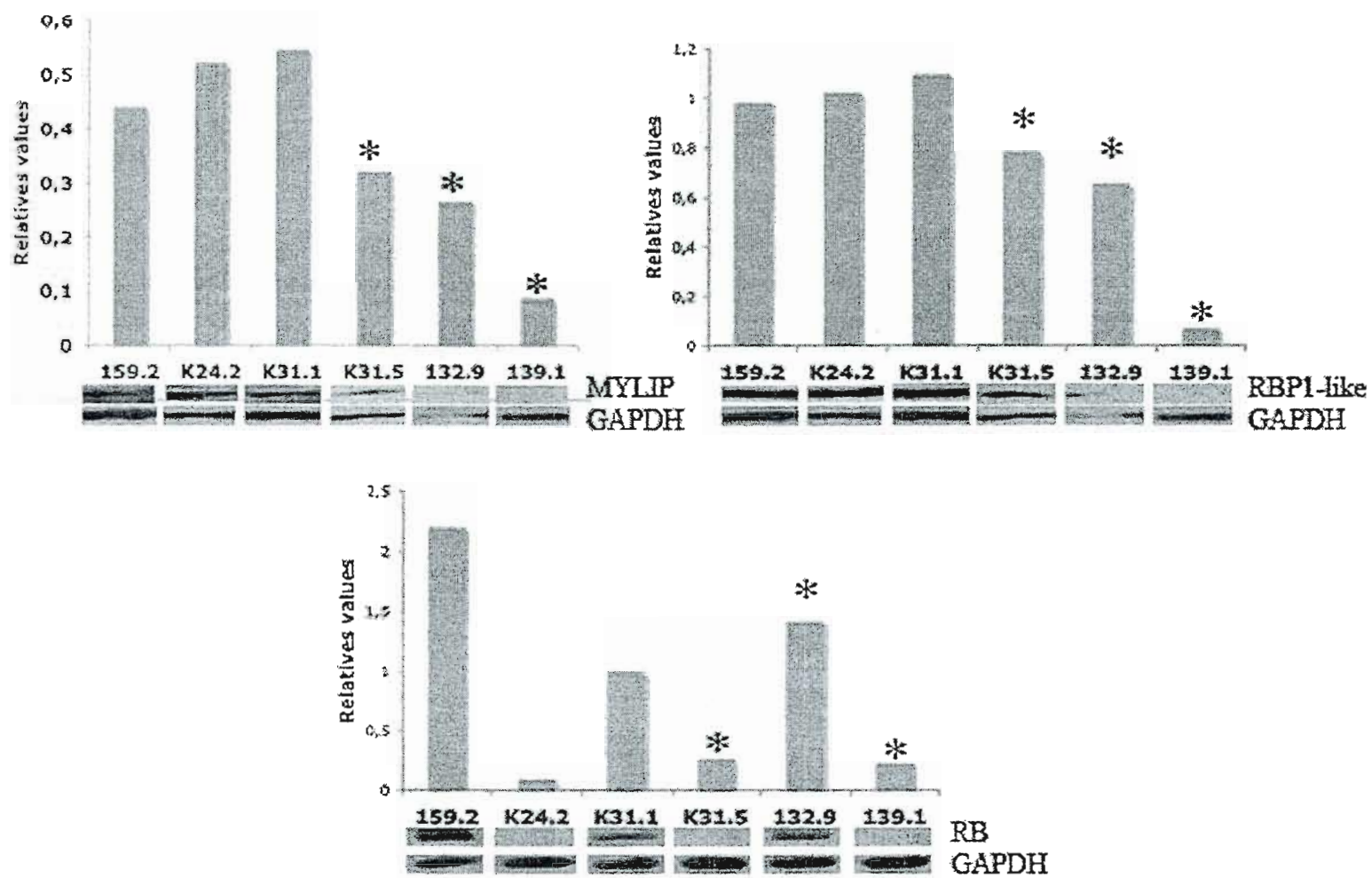


Figure 3.5C



## CHAPITRE IV

### IDENTIFICATION DES PARTENAIRES D'INTÉRACTION PROTÉIQUE DES ARN *KIS2*

#### 4.1 INTRODUCTION

Ce dernier chapitre concerne un aspect différent de la caractérisation des ARN *Kis2*, dont l'idée est venue avant la découverte du groupe *miR-106-363*.

Lorsque j'ai commencé à travailler avec le gène *Kis2*, il n'existait pas vraiment de repères quant à la caractérisation des ARNnc. En effet, peu de connaissances sont disponibles concernant par exemple les domaines structuraux ou des séquences consensus des ARNnc. Le plus souvent, l'expression des ARNnc est associée à un phénotype sans que l'on sache comment. La découverte des miARN lève une partie du mystère, puisque certains ARNnc hébergent ces petits ARN, comme c'est d'ailleurs le cas des ARN *Kis2*. L'importance des miARN a surtout apporté plus de crédit au monde des ARNnc, qui est à présent propulsé sur l'avant de la scène. Ceci devrait permettre de réels progrès quant à la caractérisation et au fonctionnement de ces derniers.

La plupart du temps, les ARN sont perçus comme une simple molécule d'ARN. Or, un ARN n'est jamais nu dans une cellule, mais toujours associé à des protéines. Les ARN couplés à leurs protéines sont appelés de façon très générale les *Ribo-Nucleo Proteins* (RNP). Tout au long de leur cycle, de leur synthèse à leur dégradation, les ARN sont donc couplés à une pléiade de protéines, que se soit lors de la transcription, l'épissage, la traduction ou le transport. En fait, les RNPs sont la forme fonctionnelle des ARN, et les protéines qui lient l'ARN sont communément appelées les *RNA Binding Proteins* (RBP), ou *heterogeneous nuclear Ribo-Nucleo Proteins* (hnRNP).

Le profil de RBPs pour un ARN donné est extrêmement dynamique (Dreyfuss *et al.* 2002). C'est notamment le cas lors de l'épissage, lorsque la séquence et la structure de l'ARN changent : des protéines se détachent alors que d'autres s'attachent. La composition des RNPs est spécifique de chaque transcrit d'une part, et dépend de la présence et de la quantité de chaque hnRNPs dans la cellule d'autre part. Il existerait donc un véritable code de RNPs, spécifique d'un transcrit donné, de son cycle, mais aussi du type cellulaire (Markovtsov *et al.* 2000, Singh *et al.* 2005).

Certains chercheurs pensent que si ce code pouvait être déchiffré, il serait quasiment plus informatif que la séquence de l'ARN même. Dans le cas d'ARN codants, la composition des RNPs ne connaît d'intérêt que lors de l'étude de la régulation post-transcriptionnelle. Mais dans le cas d'un ARNnc, c'est une porte ouverte sur la fonction de ce dernier. Prenons le cas des ARNnc TRE (*Trithorax Response Element*) chez la Drosophile (Sanchez-Elsner *et al.* 2006), qui représente une élégante élucidation du rôle d'un ARNnc. La transcription de ces ARN est initiée 30 Kpb en amont du gène Ubx (Ultrabithorax) qui est un gène dont la régulation d'expression dépend, entre autres, de facteurs épigénétiques. Les ARN TRE sont parfaitement non-codants, mais leur expression est reliée à celle de Ubx. Le lien entre TRE et Ubx a été élucidé lorsque l'interaction entre TRE et la protéine Ash1 a été montrée. Cette protéine est une histone méthyltransférase, qui active la chromatine. Dans ce cas-ci, les ARN TRE servent de point d'ancrage à Ash1, qui peut alors activer la chromatine au niveau du gène Ubx, et permettre sa transcription.

Dans le but d'élucider la fonction des ARN *Kis2*, j'ai donc voulu identifier les RBPs qui leurs sont associées. C'est en quelque sorte l'application de l'adage « Dis moi avec qui tu interagis, et je te dirais qui tu es ». Dans la partie résultats de ce chapitre, nous verrons que les protéines hnRNP A2/B1, hnRNP A3, PTB et KSRP ont été identifiées comme facteurs potentiellement associés aux ARN *Kis2*. De façon très intéressante, toutes ces protéines ont un lien avec la régulation de l'épissage. Étant donné que nous avons identifié les ARN *Kis2* comme les pri-miARN de *miR-106-363*, ces résultats n'en présentent que plus d'intérêt, notamment dans le cadre de la régulation post-transcriptionnelle de l'expression des miARN.

## MÉTHODOLOGIE GÉNÉRALE

PRODUCTION PAR TRANSCRIPTION *IN VITRO* DES ARN *KIS2*



ACCROCHAGE DES ARN À DES BILLES DE SEPHAROSE 4B



INCUBATION *IN VITRO* DES ARN COUPLÉS AUX BILLES AVEC DES  
EXTRAITS PROTÉIQUES DE CELLULES BL/VL3



ENVOI DES ÉCHANTILLONS AU CENTRE DE PROTÉOMIQUE DE  
L'UNIVERSITÉ MCGILL POUR SÉPARATION SUR GEL



SÉQUENÇAGE DES BANDES SÉLECTIONNÉES SUR GEL



CONFIRMATIONS DES INTERACTIONS ARN *KIS2*/PROTÉINES *IN VIVO*

## 4.2 MATÉRIEL ET MÉTHODES

### 4.2.1 Transcription *in vitro* (IVT)

Pour l'IVT des ARNs Kis2, les ADNc correspondants ont été clonés dans les vecteurs suivants : Kis2 t1 et Kis2 t3 dans pZero, Kis t2, t4 et t5 dans pDrive. Les vecteurs ont été digérés en grosse quantité (40 µg) en s'assurant que la digestion était totale. Kis2 t1 et t3 ont été digérés par BamH1, et Kis2 t2, t4 et t5 par Xba1 (enzymes dans le vecteur situées en 5' des ADNc). Après digestion, 1/20<sup>ème</sup> de volume de 0,5M EDTA a été ajouté, ainsi que 1/10<sup>ème</sup> de volume de 3M Na acetate et 2 volumes d'éthanol. Les échantillons ont été incubés 15min à -20°C, et centrifugés 15min à vitesse maximale dans une microcentrifugeuse. Après précipitation, l'ADN a été resuspendu dans du TE, et soumis à un traitement à la protéinase K pendant 1h à 50°C en ajoutant 0,2 µg/µL de protéinase K et 0,5 % SDS. L'ADN a ensuite été soigneusement purifié par un traitement au phénol/chloroforme, suivi d'un lavage au chloroforme et d'une précipitation à l'éthanol. Pour la transcription, 1,5 µg de chaque ADN a été utilisé selon les recommandations du manufacturier (Ambion, MEGAscript™) avec la T7 ARN polymérase 4h à 37°C. Après l'IVT, les ARNs nouvellement produits ont été nettoyés au phénol/chloroforme, et précipités à l'isopropanol. La qualité de l'ARN a été appréciée sur gel dénaturant contenant du formaldéhyde (même formule de gel que pour les Northern-blot), et la quantité déterminée par prise de D.O à 260 nm.

### 4.2.2 Fixation aux billes de Sepharose 4B activées au CNBr

Pour chaque essai, 50 µL de gel (correspondant à 14,85 mg de billes de Sepharose 4B activées au CNBr, Amersham) ont été utilisés. Les billes sont d'abord «gonflées» dans 10 mL d'HCL 1 mM (ajouté par aliquots) pendant 15 min sous agitation, suivit de 2 lavages dans du tampon potassium-phosphate 0,066 M (pH 8,3). L'ARN (10 µg) est lyophilisé au speed-vac, et repris dans 50 µL de tampon potassium-phosphate et 50% formamide (du volume final). Les ARNs sont alors incubés 10 min à 65°C, avant d'être incubés avec le gel de sépharose toute la nuit à 4°C. Le lendemain, l'ARN couplé

aux billes est lavé deux fois dans du tampon potassium-phosphate, et incubé 2h à température ambiante dans 1 mL de tampon Tris-HCl 0,1 M (pH 8,0) (blocage des groupes actifs restants). Ensuite, les billes subissent trois séries de lavages, où chaque série correspond à un lavage avec du sodium-acétate 0,1 M (pH 4,0)- 0,5 M NaCl et un lavage avec du Tris-HCl 0,1 M (pH 8,0)- 0,5 M NaCl. Les billes sont finalement rincées deux fois dans du tampon d'accrochage (250mM NaCl- 10 mM Tris-HCL (pH 7,5)- 2,5 mM MgCl<sub>2</sub>- 0,5% Triton X-100) avant d'être incubées avec les extraits protéiques.

Pour la mise au point du couplage des ARN aux billes, les ARN étaient préalablement marqués au <sup>32</sup>P (αCTP) lors de la transcription (20 μCi par échantillon). L'activité spécifique de l'ARN était évaluée au compteur à scintillation et exprimée par le rapport nombre de coups/quantité d'ARN synthétisé. Après incubation avec les billes et lavages (les surnageants étaient également conservés pour le compte), le rendement de l'opération «accrochage» aux billes était évalué avec le compteur à scintillation. Selon la quantité totale de départ, un pourcentage de fixation de l'ARN aux billes pouvait être déduit. L'ajout de formamide jusqu'à 50% du volume final m'a permis d'obtenir un taux de fixation des ARN *Kis2* aux billes de 80-90%.

#### 4.2.3 Incubation avec les extraits protéiques

Les extraits protéiques utilisés étaient des extraits nucléaires de cellules BL/VL3 préparés selon la méthode de Dignam (Dignam 1983). Les protéines obtenues (20 μg dans le tampon D: 20 mM HEPES (pH 7.9)- 20% (v/v) glycerol- 0,1 M KCl- 0,2 mM EDTA- 0,5 mM PMSF-0,5 mM DTT) sont incubées avec les billes couplées aux ARN dans un volume final de 250 μL de tampon d'accrochage (250 mM NaCl-10 mM Tris-HCl (pH 7,5)-2,5 mM MgCl<sub>2</sub>-0,5% Triton X-100), pendant 45 min à température ambiante et sous agitation. Les billes sont ensuite centrifugées et le surnageant éliminé. Les échantillons sont alors soumis à une série de quatre lavages, dont le premier consiste à une incubation de 10 min à 4°C dans du tampon d'accrochage. Le deuxième lavage est fait dans du tampon d'accrochage 2mg/mL héparine, 10 min 4°C. Les troisième et quatrième lavages, de 10 min chacun, sont fait dans du tampon d'accrochage sans héparine, à 4°C.

#### 4.2.4 Séparation sur gel et séquençage des protéines

Après incubation avec les extraits protéiques, les échantillons ont été directement mis sur glace, et apportés au centre de protéomique de l'Université McGill. Les protéines ont été séparées sur un gradient de gel de polyacrylamide, coloré par la suite au nitrate d'argent. Dans cette expérience, les billes de sépharose 4B seules incubées avec les extraits protéiques servent de contrôle négatif. Seules les bandes protéiques dont aucune trace n'était décelable dans le contrôle négatif ont été choisies pour séquençage par spectroscopie de masse. Chaque bande choisie a été séquencée pour *Kis2* t1 et *Kis2* t3.

#### 4.2.5 RNA Immunoprecipitation Assay (RIP)

Dans un premier temps, 100 µL de billes de sépharose couplées à la protéine G (Santa Cruz) sont incubées avec 30 µg d'anticorps, dans 250 µL (volume final) de tampon NET2 (50 mM Tris-HCl (pH 7,4)-150 mM NaCl-1 mM MgCl<sub>2</sub>-0,05% NP40) 1h (ou toute la nuit) à 4°C. Les billes couplées aux anticorps sont ensuite lavées 5 fois dans du tampon NET2, et incubées toute la nuit à 4°C avec 200 µg d'extraits protéiques dans 300 µL final de tampon NET2 contenant un inhibiteur de protéase (Roche Applied Bioscience) et 100 U d'inhibiteur de RNase (Amersham). Des billes non couplées aux anticorps sont utilisées comme contrôle négatif. Les extraits protéiques utilisés sont toujours des extraits nucléaires de cellules BL/VL3 purifiés par la méthode de Dignam, (et sont donc en solution dans du tampon D). Le lendemain, les billes sont lavées 5 fois avec du tampon NET2, le premier surnageant étant conservé. Les billes et le premier surnageant sont ensuite soumis à une extraction d'ARN au TRIzol (Invitrogen). La totalité de l'ARN extrait est utilisée pour la RT-PCR subséquente.

La RT (Omniscript, Qiagen) est initiée avec un oligonucléotide spécifique de *Kis2* (0,5 µM final) (Oligo B, 5' GACATGGCCAAGGCATACTCCAGGAGTGTAAACC 3') selon les recommandations du manufacturier. Pour la PCR (Qiagen), la totalité de la RT est employée dans une réaction de 50 µL selon le protocole, et 0,2 µM final des oligonucléotides A (5' GGCCGCTAAGGTGTCTTGAAGACCTAAGCGCG 3') et B. Le cycle de PCR est le suivant: 90°C 2 min suivit de 90°C 30 sec-63°C 30 sec-72°C 1 min 30 sec répété 35 fois. La moitié de la réaction de PCR est ensuite visualisée sur gel d'agarose 1%.

#### 4.2.6 Western-Blot

Les Western-blot pour visualiser les protéines PTB, KSRP, hnRNP A2/B1 et A3 ont été réalisés comme suit: les extraits nucléaires de cellules BL/VL3 ou NIH/3T3 ont été préparés selon la méthode de Dignam comme décrit précédemment, et 20 µg de chaque extrait ont été mélangés avec du bleu de chargement 6X (0.35M Tris-HCl [pH 6.8], 10 % SDS, 36 % glycerol, 5 % β-mercaptoethanol, 0.012 bromophenol blue), puis bouillis 5 min à 100°C. Les protéines ont été séparées sur un gel d'électrophorèse de SDS-polyacrylamide 0,8 %, et transférées par électrophorèse sur une membrane de polyvinylidene difluoride (Immobilon-P, Millipore). Les membranes ont été bloquées dans du PBS 1X-0.2% Tween 20-5% lait en poudre écrémé pendant 1 h à température ambiante, et incubées toute la nuit à 4°C avec les anticorps dilués dans du PBS-Tween 5 % BSA. Les dilutions utilisées pour les anticorps furent les suivantes: hnRNP A2/B1 (Santa Cruz) 1/500, hnRNP A3 (Santa Cruz) 1/500, PTB (gracieusement offert par Douglas L. Black, Department of Microbiology, Immunology, et Molecular Genetics, University of California, Los Angeles, Los Angeles, California 90095, USA) 1/1500, KSRP (gracieusement offert par Douglas L. Black) 1/3000. Les membranes ont été lavées dans du PBS-Tween 6 fois 10 min, et incubées 1h dans du PBS-Tween-lait avec les anticorps secondaires couplés à la peroxydase (anti-souris (GE Healthcare) 1/10 000 pour hnRNP A2 et KSRP, anti-lapin (GE Healthcare) 1/20 000 pour PTB, et anti chèvre (Zymed) 1/5000 pour hnRNP A3). Le signal a été révélé avec Immobilon Western (Millipore) selon le protocole recommandé, suivi d'une exposition sur film (Hyperfilm ECL, GE healthcare).

#### 4.2.7 Transfection ADNc *Kis2*

Pour l'expression des ARN *Kis2* dans les cellules NIH/3T3, les ADNc correspondant à chacun des transcrits ont été clonés dans le vecteur pCMV5. Les ADN ont été transfectés dans des cellules NIH/3T3 en utilisant l'agent de transfection Lipofectamine 2000 (Invitrogen) selon les recommandations du manufacturier. Les cellules (800 000) ont étéensemencées dans des pétris 60 mm, et transfectées 24h après avec 20 µL de lipofectamine et 8 µg d'ADN. Les cellules ont été récoltées et lysées

directement dans le TRIzol 48h après la transfection. L'ARN extrait a ensuite été utilisé en Northern-blot pour visualiser les ARN *Kis2*.

#### 4.2.8 Northern-blot

Les ARN totaux (15 µg) ont été séparés sur gel d'agarose 1% contenant du formaldéhyde, selon le protocole déjà décrit (Birren *et al.* 1998), puis transférés par capillarité sur membrane de nylon. La membrane a été préhybridée pendant 1h à 42°C dans 50% formamide- 5X SSPE (3 M NaCl, 20 mM EDTA, 20 mM sodium phosphate pH 6.8)- 5X Denhardt's- 0.5% SDS- 0.1 mg/ml ADN dénaturé de sperme de saumon. L'hybridation a été effectuée toute la nuit à 42°C dans le même tampon contenant 10% de sulfate de dextran. La sonde spécifique de *Kis2* (sonde RTB3) a été marquée par la méthode de «random primer extension» en utilisant des oligohexamères (GE Healthcare), et approximativement  $2 \times 10^6$  cpm/ml ont été utilisés. Les membranes ont été lavées 10 min à 42°C dans du 2X SSC (1X SSC = 0.15 M NaCl- 0.015 M sodium citrate), 30 min à 42°C dans du 2X SSC-0.1% SDS, et finalement 15 min à 65°C dans du 0.1X SSC- 0.1% SDS. Les membranes ont été révélées en utilisant un phosphorimager (BioRad).



## 4.3 RÉSULTATS

### 4.3.1 Production des ARN *Kis2* par transcription *in vitro* et fixation aux billes de CnBr

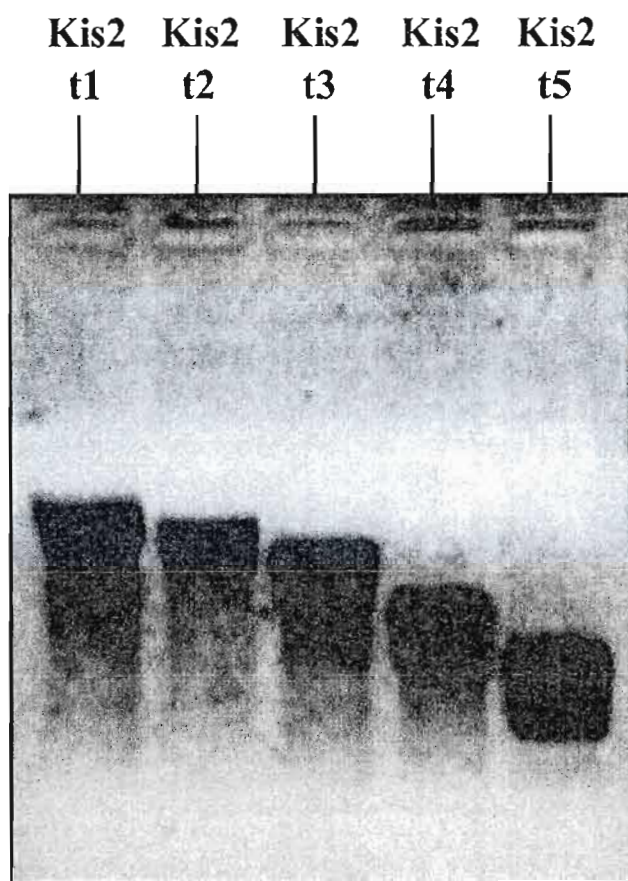
La première étape de ce travail a consisté en la production par transcription *in vitro* des ARN *Kis2*. Il s'agit donc ici de la partie du pri-miARN de *miR-106-363* qui a été initialement identifiée en Northern-blot dans l'article du chapitre II (Landais *et al.* 2005). C'est également la partie en rouge des pri-miARN représentés dans la figure 3.0. Rappelons qu'à l'époque où j'ai commencé ce travail, je ne connaissais pas l'existence de *miR-106-363*. La transcription *in vitro* des ARN *Kis2* a permis d'obtenir des ARN de bonne qualité, dans des quantités satisfaisantes (Fig. 4.1).

Un bon pourcentage de fixation de chacun des ARN aux billes de Sepharose a pu être obtenu (autours de 90%), moyennant des mises au point des conditions de couplage. Notamment, l'augmentation de la concentration de formamide s'est avérée décisive, ce qui traduit probablement l'existence de structures secondaires complexes au niveau des ARN *Kis2*.

### 4.3.2 Identification de protéines interagissant avec les ARN *Kis2*

Afin d'identifier des protéines qui interagissent avec les ARN *Kis2*, j'ai choisi d'utiliser des extraits protéiques issus de la lignée cellulaire BL/VL3. Ces cellules, qui proviennent d'un lymphome T de souris, ont en effet la propriété de surexprimer légèrement les ARN *Kis2* (Landais *et al.* 2005), et elles représentent de plus une source illimitée de matériel puisque nous la possédons dans notre laboratoire. Les extraits protéiques nucléaires ont été préférés aux extraits totaux, car les ARN *Kis2* sont majoritairement nucléaires.

Le problème du choix d'un contrôle négatif approprié s'est posé : quel type d'ARN «neutre» pouvait-on utiliser pour remplir le rôle de contrôle négatif? Si des protéines interagissent avec un ARN, quel qu'il soit, ça ne veut pas dire qu'elles n'interagissent pas non plus avec les ARN *Kis2*. Pour cette raison, et dans l'optique



**Figure 4.1** Transcription in vitro des ARN *Kis2*. Dépôt de 1  $\mu$ L sur gel d'agarose 1% contenant du formaldéhyde.

d'une confirmation des résultats *in vivo* par la suite, j'ai choisi d'utiliser les billes seules comme contrôle négatif.

Une fois tous les paramètres optimisés, l'expérience a été conduite selon les conditions sélectionnées, et les échantillons ont été soumis au centre de protéomique de l'Université McGill. Dans un premier temps, les protéines ont été séparées sur gel (Fig. 4.2). Les cinq ARN *Kis2* incubés avec les extraits protéiques sont donc comparés avec les billes seules incubées avec les mêmes extraits.

Plusieurs bandes étaient présentes dans les pistes correspondant aux ARN *Kis2* et non présentes dans le contrôle négatif. Parmi ces bandes, celles dont aucune trace n'était décelable dans le contrôle négatif ont été préférées, et sur cette base, six bandes ont été sélectionnées (Fig. 4.2). Pour la bande 1, les résultats indiquent clairement qu'il s'agit de la protéine hnRNP A2/B1 (Tableau 4.1), pour laquelle A2 et B1 sont deux variants d'épissage qui diffèrent par douze acides aminés de plus pour B1 (Burd *et al.* 1989, Kozu *et al.* 1995).

**Tableau 4.1 Résumé du séquençage par spectroscopie de masse des bandes sélectionnées sur gel**

GEL	SÉQUENÇAGE	PEPTIDES*	SCORE	MASSE (kDa)
Bande 1	hnRNP A2	17/35	768	36
Bande 2	hnRNP B1	12/19	601	37,5
	hnRNP A3	15/21	692	39,8
Bande 3	hnRNP A3	12/18	575	39,8
Bande 4	Kératine			
Bande 5	PSF	3/4	217	75,5
	KSRP	2/4	192	76
Bande 6	KSRP	20/27	1074	76

\* Cette donnée concerne l'ensemble des peptides qui ont été identifiés dans cet échantillon et qui ont servi à identifier la protéine correspondante. Par exemple, la donnée «17/35» signifie que sur 35 peptides identifiés dans l'échantillon de gel, 17 appartiennent significativement à la protéine déduite.

Pour la bande 2, les résultats indiquent avec un score similaire qu'il s'agit soit de hnRNP A2/B1, soit de hnRNP A3. Cependant, la présence du peptide «KTLETVP LERKK» spécifique de hnRNP B1, et absent de hnRNP A3, nous amène à penser qu'il s'agit plutôt de hnRNP B1. En conséquence, la bande 1 correspondrait plutôt à hnRNP A2, tandis que la bande 3 est identifiée comme étant hnRNP A3. La bande 4 s'est avérée être la kératine, et très probablement une contamination. La bande 5 se solde par un résultat équivoque mêlant PSF (*PTB associated splicing factor*) et KSRP (*K-homology Splicing Regulatory Protein*). Seul un petit nombre de peptide est détecté, avec un score très similaire pour les deux protéines. La taille estimée pour les deux protéines semble également correspondre à celle observée sur gel (autour de 75 kDa). Pour finir, la bande 6 correspondrait clairement à KSRP, bien que la taille sur le gel soit plus haute que la masse prédite pour cette protéine. Cependant, cette fois-ci, un grand nombre de peptides correspondant à cette protéine sont présents, ce qui lève le doute quant à la nature de cette bande. Des modifications post-traductionnelles peuvent être à l'origine de la différence entre la taille observée et théorique de KSRP.

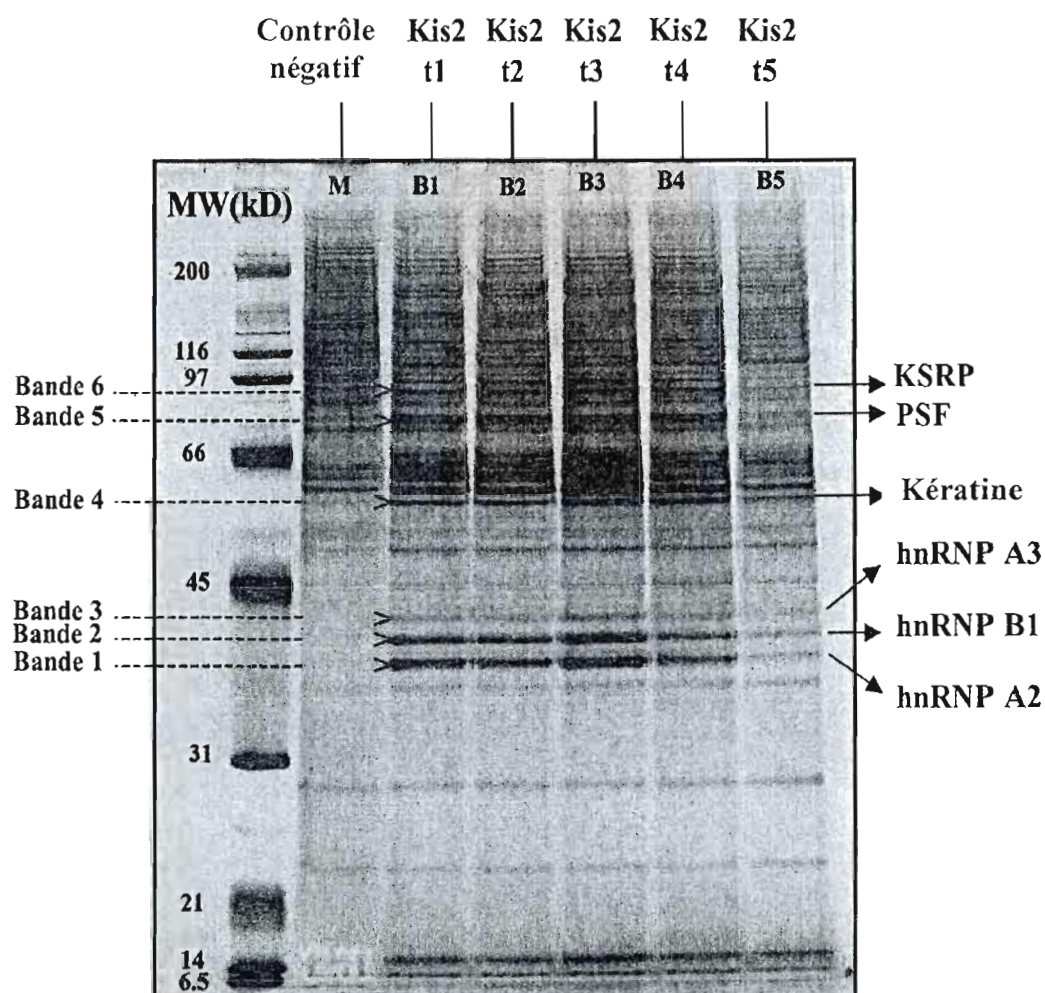
En conclusion, les protéines hnRNP A2, hnRNP B1, hnRNP A3 et KSRP ont été clairement identifiées par séquençage. Le résultat pour PSF reste moins certain.

#### 4.3.3 Intérêt des protéines hnRNP A2/B1, hnRNP A3, PTB et KSRP

Le point commun entre hnRNP A2/B1, hnRNP A3, KSRP et PSF est leur rapport avec l'épissage.

KSRP est un facteur qui a été premièrement rapporté pour favoriser l'épissage de l'exon N1 de c-src, via un ISE (Min *et al.* 1997). Cette protéine possède quatre domaines de liaison à l'ARN de type KH. La littérature ne compte pas d'autres rapports impliquant KSRP dans l'épissage, mais par contre de plus en plus de publications font état du rôle de KSRP dans la dégradation de l'ARN via les AREs (AU Rich Elements), souvent situés dans les régions 3' non codantes des ARNm (Chou *et al.* 2006, Gherzi *et al.* 2004).

hnRNP A2/B1 est également une protéine qui lie l'ARN grâce à deux domaines de type RRM. Cette protéine a diverses fonctions, et est notamment connue pour interagir avec les télomères (McKay *et al.* 1992, Ishikawa *et al.* 1993, Moran-Jones *et al.* 2005). Cette protéine peut apparemment reconnaître différents motifs dans l'ARN, comme



**Figure 4.2 Séparation des protéines sur gel de polyacrylamide (gradient).**

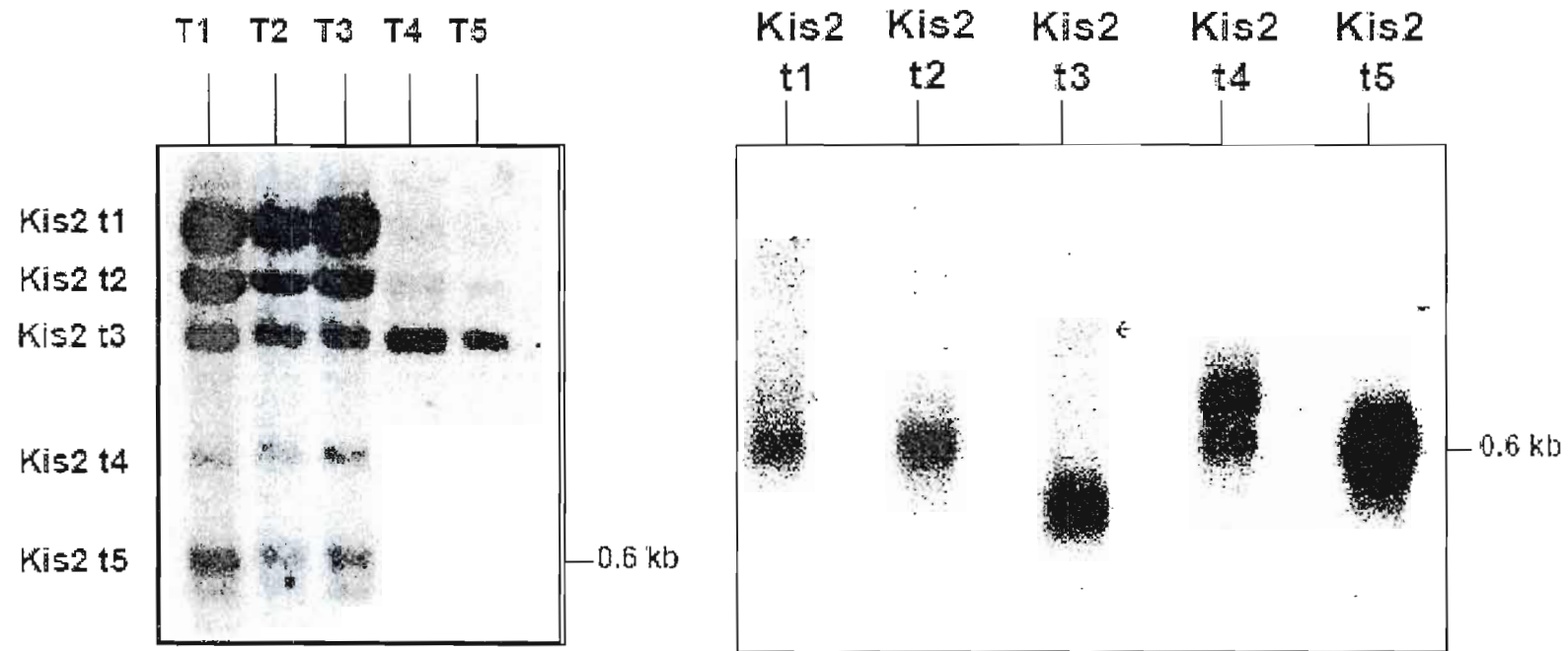
MW, Molecular Weight; kD, kilo Dalton. Le contrôle négatif correspond aux billes de CNBr seules incubées avec les extraits protéiques de cellules BLVL3. Chaque transcrit *Kis2* a été fixé aux billes et incubé avec ces extraits. Les bandes indiquées (1-6) ont été excisées et les protéines identifiées par spectroscopie de masse.

un ARE dans le 3'UTR du transporteur de glucose GLUT1 (Griffin *et al.* 2004), l'élément A2RE/A2RE11 de transport cytoplasmique des ARN (Munro *et al.* 1999, Landsberg *et al.* 2006), ou un motif proche de la séquence consensus du site 3' d'épissage (Ishikawa *et al.* 1993). Bien que le rôle de cette protéine soit plus documenté en ce qui concerne le transport des ARN, plusieurs rapports font état du rôle de hnRNP A2/B1 dans la sélection de sites d'épissage 5' et 3' (Mayeda *et al.* 1994, Caputi *et al.* 1999, Bilodeau *et al.* 2001, Hutchinson *et al.* 2002).

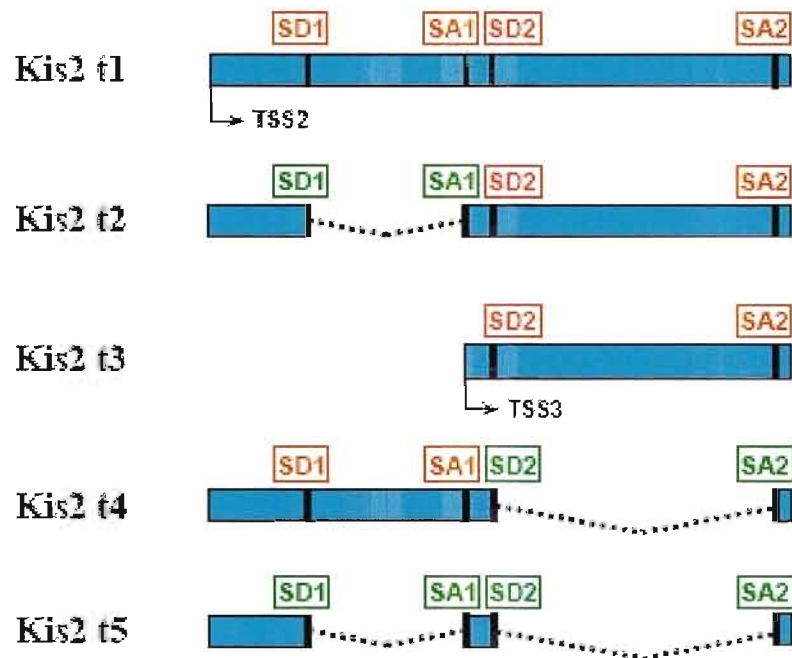
En ce qui concerne hnRNP A3, peu de renseignements sont disponibles quant à sa fonction. Cette protéine a cependant été impliquée dans le transport des ARN, par son interaction avec des motifs de type A2RE, et pourrait interagir avec hnRNP A2/B1 (Ma *et al.* 2002). En effet, plusieurs publications rapportent la présence lors de l'épissage de plusieurs hnRNPs dont hnRNP A2/B1 et hnRNP A3 (Rappsilber *et al.* 2002, Zhou *et al.* 2002, Kasyapa *et al.* 2005).

PSF est directement impliquée dans l'épissage et a été montrée comme essentielle à l'assemblage du spliceosome (Gozani *et al.* 1994, Patton *et al.* 1993). Elle a initialement été identifiée comme partenaire de la protéine PTB (Patton *et al.* 1993), et comme cette dernière, PSF se fixe à la séquence riche en pyrimidines située en 3' des sites d'épissage. Plus récemment, PSF a été impliquée en tant qu'activateur dans le couplage de l'épissage à la transcription (Rosonina *et al.* 2005).

Même si PTB ne fait pas partie des protéines qui ont été séquencées lors de notre expérience, elle n'en est pas moins intéressante. Non seulement elle est souvent associée à PSF, mais elle peut également l'être avec hnRNP A2/B1 et A3 (Rappsilber *et al.* 2002, Zhou *et al.* 2002, Kasyapa *et al.* 2005) ainsi que KSRP (Markovstov *et al.* 2000). PTB participe à la régulation de l'épissage, notamment en tant qu'inhibiteur (voir Chapitre I, 1.3.2.3). Pour ces raisons, et également parce que nous pouvions facilement nous procurer des anticorps dirigés contre PTB, nous avons inclus cette protéine dans la suite de notre étude. Par contre, PSF ne sera pour l'instant pas investiguée. L'ambiguïté du séquençage et son caractère essentiel/ubiquitaire à l'épissage en font une candidate moins intéressante dans un premier temps.



**Figure 4.3 Comparaison entre les ARN *Kis2* observés dans les tumeurs et les ARN *Kis2* dans les cellules NIH/3T3.** (A) Northern-blot spécifique des ARN *Kis2* dans les tumeurs (Landais *et al.* 2005). (B) Les ADNc des ARN *Kis2* ont été transfectés dans des cellules NIH/3T3. Seul le transcrit de 0,6 kb est visible. Notez que pour *Kis2 t3*, l'épissage donne lieu à un fragment plus petit (Landais *et al.* 2005).



**Figure 4.4** Combinaison d'inhibition/ activation d'épissage à l'origine des transcrits *Kis2*. Les différents transcrits *Kis2* résultent d'un ensemble d'inhibition (en rouge) et d'activation (vert) des sites donneurs (SD) et accepteurs (SA) d'épissage.



#### 4.3.4 Le profil d'épissage des ARN *Kis2* observé dans les tumeurs diffère de celui observé dans les cellules NIH/3T3.

En plus d'un profil de littérature pour ces protéines orientées vers la régulation de l'épissage, et du profil complexe d'épissage des ARN *Kis2*, un indice expérimental vient s'ajouter. En effet, lorsque les ADNc correspondant aux cinq ARN *Kis2* surexprimés dans les tumeurs de souris étaient transfectés dans des cellules NIH/3T3, seul le plus petit des transcrits *Kis2* (*Kis2* t5) était observable en Northern-blot (Fig. 4.3). Les plus grandes espèces d'ARN *Kis2* (*Kis2* t1 à t3) étaient totalement épissées pour donner le plus petit transcrit. Or, dans les tumeurs réarrangées pour *Kis2* surexprimant les ARN, c'est le contraire qui est observé : ce sont les transcrits non épissés qui sont majoritairement représentés. Ceci nous amène donc à penser que des facteurs protéiques, impliqués dans l'épissage et/ou la (dé)stabilisation des ARN *Kis2*, sont présents de façon différentielle entre les cellules tumorales et les cellules NIH/3T3. La production des différents transcrits *Kis2* dans les tumeurs résulte d'une combinaison d'inhibition/activation des différents sites d'épissage (Fig 4.4), contrôlée par des protéines régulatrices. Comme les ARN *Kis2* non épissés (t1 à t3) sont majoritairement représentés dans les tumeurs, l'identification des facteurs responsables de leur expression s'avère donc de premier intérêt. À ce titre, l'interaction possible des ARN *Kis2* avec les protéines hnRNP A2/B1, hnRNP A3, KSRP et PTB mérite d'être investiguée.

#### 4.3.5 Analyse d'expression des protéines hnRNP A2/B1, hnRNP A3, PTB et KSRP dans les cellules BL/VL3 et NIH/3T3

Pour savoir si la différence d'épissage des ARN *Kis2* observée entre tumeurs et cellules NIH/3T3 pouvait être due à une différence d'expression de KSRP, PTB, hnRNP A2/B1 et A3, j'ai vérifié leur niveau d'expression en Western-blot. Pour cela, j'ai comparé l'expression de ces protéines dans les cellules NIH/3T3 par rapport à celle des cellules BL/VL3, puisque ces dernières exhibent le même profil d'épissage pour les ARN *Kis2* que dans les tumeurs. Si on se base sur les résultats de cette seule expérience, alors les résultats indiquent effectivement des différences entre les deux types cellulaires (Fig. 4.5). Par exemple, le niveau d'expression de hnRNP A3 apparaît nettement plus élevé

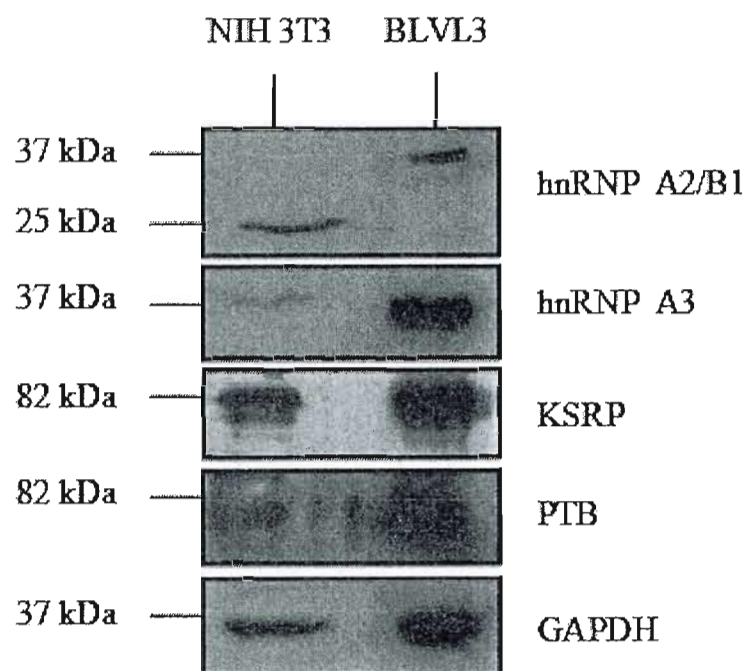


Figure 4.5 Western-blot de comparaison entre cellules NIH/3T3 et BL/VL3 pour les protéines hnRNP A2/B1, hnRNP A3, KSRP et PTB. GAPDH sert de contrôle de chargement.

dans les cellules BL/VL3 que dans les NIH/3T3. Pour hnRNP A2/B1, le niveau d'expression semble au contraire plus élevé dans les NIH/3T3 que dans BL/VL3. On peut également remarquer une différence pour la taille: le poids moléculaire de hnRNP A2/B1 apparaît plus petit ( $\approx 25$  kDa) dans les cellules NIH/3T3. Curieusement, ceci n'est observable qu'avec les extraits nucléaires de ces cellules, et non avec les extraits totaux (non montré). Ce fragment d'à peu près 25 kDa a déjà été décrit comme étant le résultat d'une dégradation, et non comme un nouvel isoforme (Williams 1985). La dégradation n'est toutefois pas observable dans les extraits nucléaires des BL/VL3. PTB semble également plus abondante dans les cellules BL/VL3 que dans NIH/3T3. En ce qui concerne KSRP, il n'y a pas vraiment de différence notable entre les deux types cellulaires. À noter cependant que la taille de PTB est prédite autour de 52 kDa chez la souris, et que la taille observée en western-blot pour les deux types de cellules se situe autour des 82 kDa. Ceci nous indique l'existence de modification post-traductionnelle de cette protéine.

En vertu de ces résultats, les différences de niveau d'expression de ces protéines entre cellules NIH/3T3 et BL/VL3 se situeraient au niveau de hnRNP A3 et PTB (surexprimé dans BL/VL3), ainsi que hnRNP A2/B1 (sous-exprimé dans BL/VL3).

#### **4.3.6 Les problèmes de confirmation des interactions des ARN *Kis2* avec les protéines**

La confirmation des résultats obtenus par spectroscopie de masse est une étape cruciale pour qu'ils puissent être validés. Pour cela, deux choix s'offraient à moi : une expérience de RNA *band shift* (*in vitro*), ou une expérience d'immunoprécipitation d'ARN (*in vivo*). Les deux expériences peuvent s'avérer coûteuses en temps de mise au point et surtout d'anticorps spécifiques des protéines testées. Pour cette raison, j'ai choisi de concentrer mes efforts sur l'immunoprécipitation d'ARN, puisque c'est une expérience *in vivo* dont le résultat aurait plus de valeur.

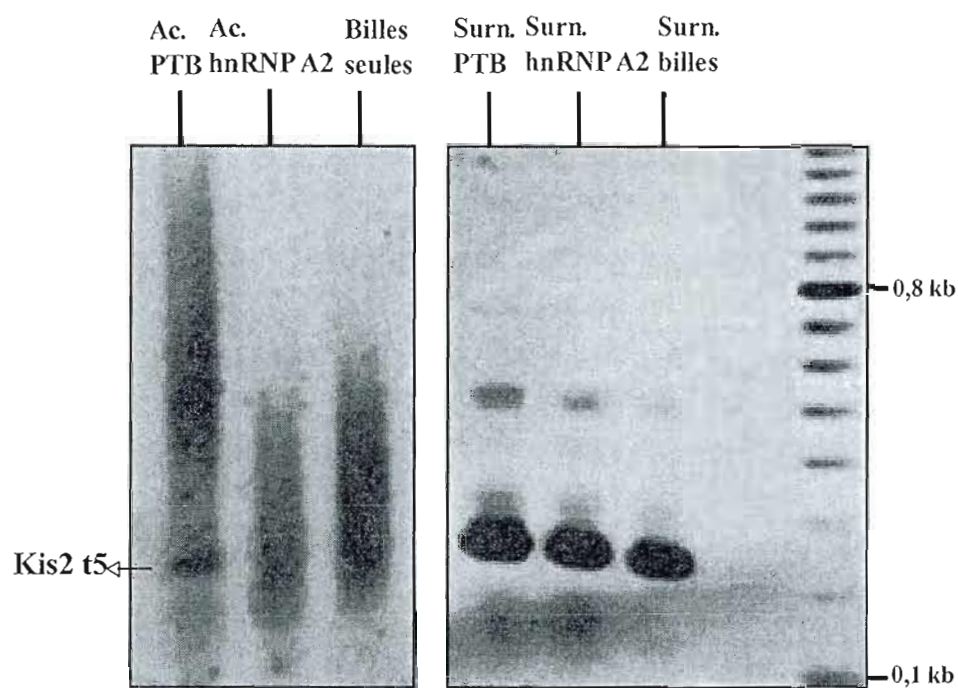
Pour réaliser l'immunoprécipitation d'ARN (RIP), je me suis confrontée à plusieurs problèmes. Le premier problème, majeur de surcroît, fut l'impossibilité de purifier des RNPs contenant les plus grands transcrits *Kis2* intacts. Les extraits protéiques ont été préparés à partir de cellules BL/VL3, puisqu'elles surexpriment naturellement les ARN *Kis2*. La préparation d'extraits nucléaires a été préférée aux extraits protéiques

totaux, puisque les ARN *Kis2* sont nucléaires. De plus, l'utilisation d'extraits nucléaires évitait la saturation inutile des anticorps par les protéines cytoplasmiques (les protéines testées sont également présentes dans le cytoplasme). À l'issue des extractions, les grands transcrits *Kis2* n'ont jamais pu être détectés dans des extraits protéiques nucléaires de BL/VL3, et seul *Kis2* t5 était détectable. Or, l'hypothèse majeure de mon travail concernait l'inhibition de l'épissage des grands transcrits de *Kis2*.

Les résultats de spectroscopie de masse montraient cependant la présence des protéines sélectionnées avec *Kis2* t5. J'ai donc au moins voulu confirmer ceci par RIP sur les extraits protéiques nucléaires de BL/VL3. L'expérience a été dans un premier temps essayée avec les anticorps spécifiques de PTB et hnRNP A2/B1, et un signal correspondant à *Kis2* t5 fut effectivement détecté pour le RIP conduit avec les anticorps anti-PTB (Fig 4.6). Par contre, aucun signal n'était décelable dans l'expérience spécifique de hnRNP A2. Un test devait plus tard me confirmer que les anticorps anti-hnRNP A2 vendus par la compagnie Santa-Cruz ne fonctionnaient pas en immunoprécipitation (pourtant garantis par la compagnie). Le résultat négatif obtenu pour hnRNP A2 était donc dû en premier lieu à une absence d'immunoprécipitation. Par la suite, un chercheur travaillant de routine avec hnRNP A2 devait me confirmer que les anticorps commerciaux spécifiques de cette protéine ne fonctionnaient effectivement pas en immunoprécipitation. J'ai donc obtenu de précieux anticorps anti-hnRNP A2 «home-made», que je n'ose pour l'instant utiliser tant que l'expérience de RIP ne sera pas exactement paramétrée.

Ceci m'amène à mon deuxième problème : les anticorps. En effet, le RIP est une expérience très coûteuse en anticorps. Pour les anticorps commerciaux, ce ne serait qu'une question de coût, et non de disponibilité. Or, trois des anticorps sur les quatre avec lesquels je travaille ne sont pas commerciaux. D'ailleurs, après avoir essayé deux fois le RIP avec les anticorps spécifiques de PTB, je n'en ai malheureusement plus.

Cependant, à défaut d'avoir une confirmation d'interaction avec les grands transcrits *Kis2*, le signal détecté pour *Kis2* t5 lors du RIP spécifique de PTB est encourageant, et nécessite l'obtention d'anticorps anti-PTB afin que l'expérience puisse être répétée.



**Figure 4.6 Résultats du RIP spécifique de PTB et hnRNP A2/B1.** Après PCR, 20  $\mu$ L d'amplifiat ont été déposés sur gel d'agarose. Les pistes «Ac.» correspondent aux extractions d'ARN, après immunoprécipitation, sur les billes couplées avec anticorps spécifiques de chaque protéine. Les billes seules (sans anticorps) ont servi de contrôle négatif. Les pistes «Surn.» correspondent aux extractions d'ARN sur les surnageants de chaque immunoprécipitation.

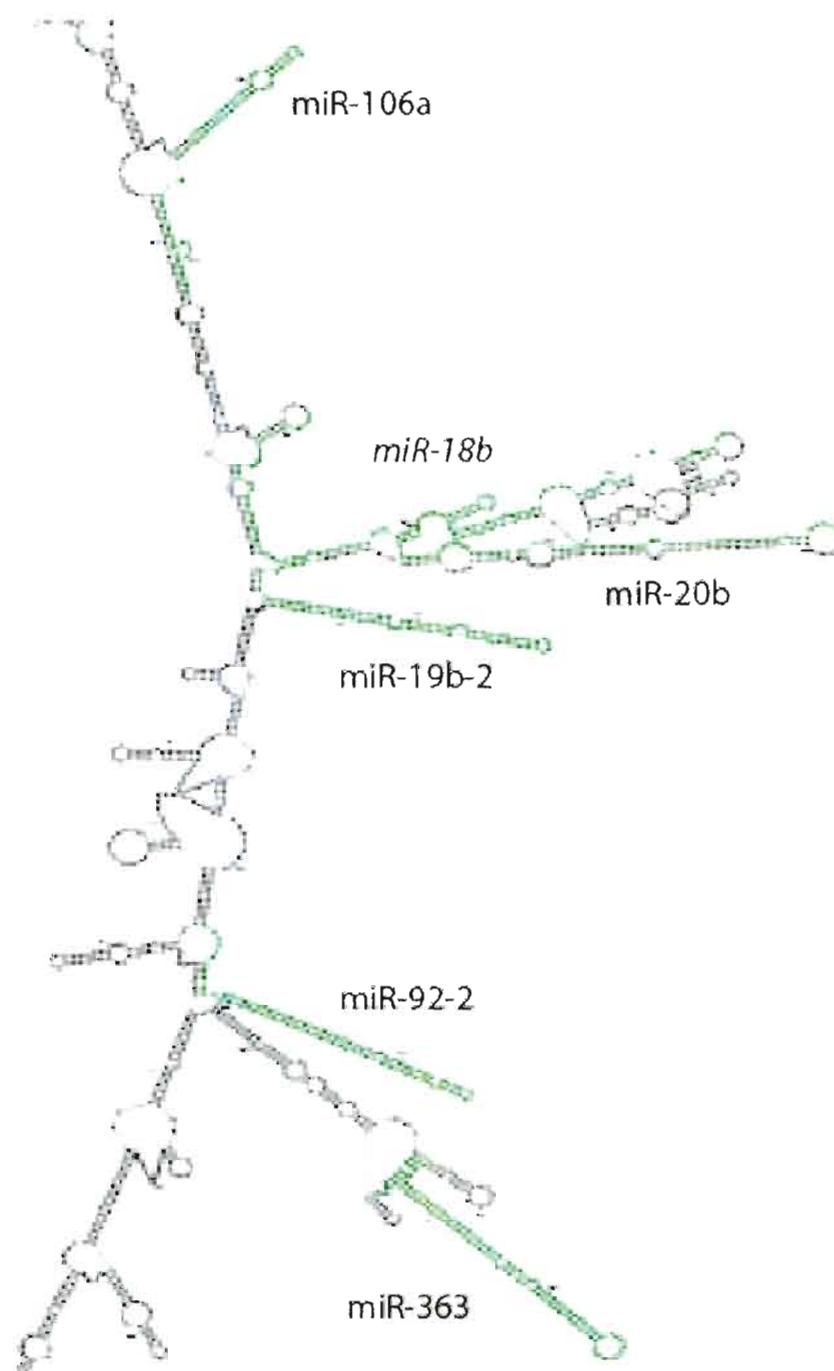
#### 4.3.7 Implication pour le groupe *miR-106-363*

La découverte que les ARN *Kis2* étaient en fait les pri-miARN du groupe *miR-106-363* a rendu les partenaires d'interaction protéiques des ARN *Kis2* encore plus attrayants. L'hypothèse qui se profilait concernait un rôle de ces protéines au niveau d'une éventuelle régulation post-transcriptionnelle de la production des miARN. Cette hypothèse était renforcée par d'autres indices. Premièrement, l'existence d'un profil complexe d'épissage des pri-miARN *Kis2* apparaît intrigante, et n'est certainement pas le fruit du hasard ou d'un «artéfact» du système d'épissage. La question de la signification de tous ces transcrits s'est donc posée depuis le départ.

Deuxièmement, tous les miARN du groupe *miR-106-363* ne sont pas produits en même temps, ce qui laisse supposer une régulation post-transcriptionnelle de leur expression. Enfin, lorsque la structure secondaire des différents pri-miARN est analysée à l'aide de programme Mfold (<http://www.bioinfo.rpi.edu/applications/mfold/rna/form1>), cette structure n'est évidemment pas la même d'un transcrit à l'autre. Si on compare les différentes structures entre elles (pri-miARN complets correspondants aux transcrits *Kis2* t1, t2, t3, t4, t5 et t6) (Fig. 4.7 à 4.12), on peut voir que les tiges boucles correspondant aux futurs pre-miARN sont invariablement présentes, mis à part pour miR-18b et miR-106a. La tige-boucle de miR-18b est par exemple visible pour les transcrits *Kis2* t4, t5 et t6 (Fig. 4.10 à 4.12), mais pas pour les transcrits *Kis2* t1, t2 et t3 (Fig. 4.7 à 4.9). La tige-boucle de miR-106a n'apparaît pas dans la prédiction de structure de *Kis2* t6 (Fig. 4.12). En ce qui concerne miR-20b, la tige-boucle prédite par Mfold n'est étrangement pas la même que celle des banques de données, et ce pour tous les transcrits (Fig. 4.7 à 4.12). Enfin de façon intéressante, si les tiges-boucles des miARN ne varient pas (à part pour miR-18b et miR-106a), le reste de la structure secondaire pour chaque transcrit présente des variations (Fig. 4.7 à 4.12).

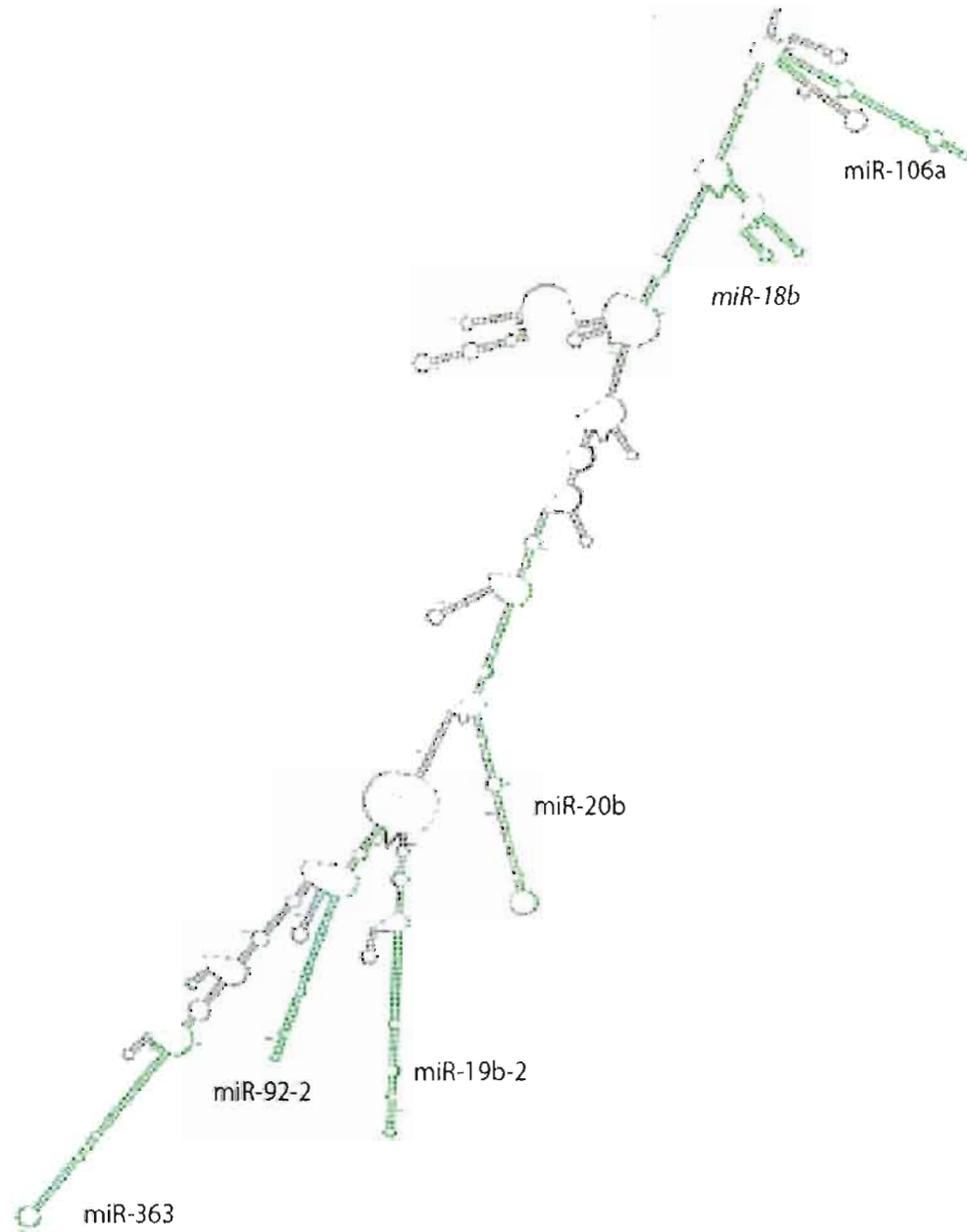
Ainsi, en fonction des différentes structures secondaires des ARN *Kis2*, la production de miARN pourrait ne pas être la même. Ceci serait dû soit à un profil différent de tiges-boucles des miARN (cas de miR-18b ou miR-106a par exemple), soit à un recrutement différent de protéines régulatrices du *processing* du pri-miARN. Les protéines qui régulent l'épissage de ces transcrits seraient donc directement impliquées dans la régulation de la production des miARN du groupe *miR-106-363*. Ceci ne

s'appliquerait pas juste au pri-miRNA de *miR-106-363*, car de nombreux autres pri-miRNAs exhibent aussi des transcrits alternatifs.

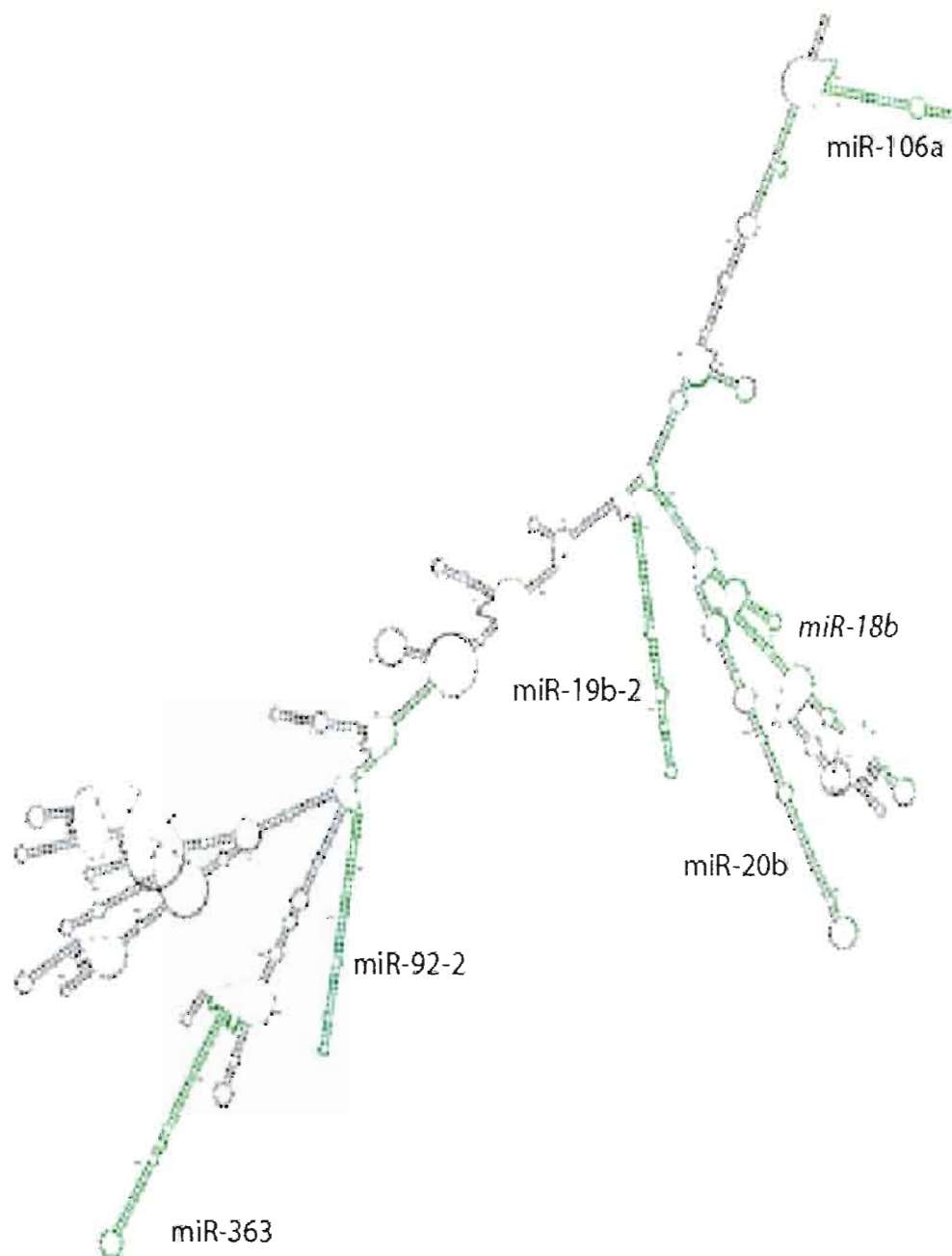


**Figure 4.7 Prédiction de la structure de l'ARN *Kis2* t1**, du pri-miARN complet correspondant au transcrit *Kis2* t1. Les portions en vert correspondent aux tiges-boucles des miARN. Ici la tige-boucle de miR-18b n'est pas apparente, et notez celle de miR-20b qui ne correspond pas à la séquence connue (représenté par le mélange de deux couleurs). Seule la partie avec les miARN est représentée pour plus de clarté.

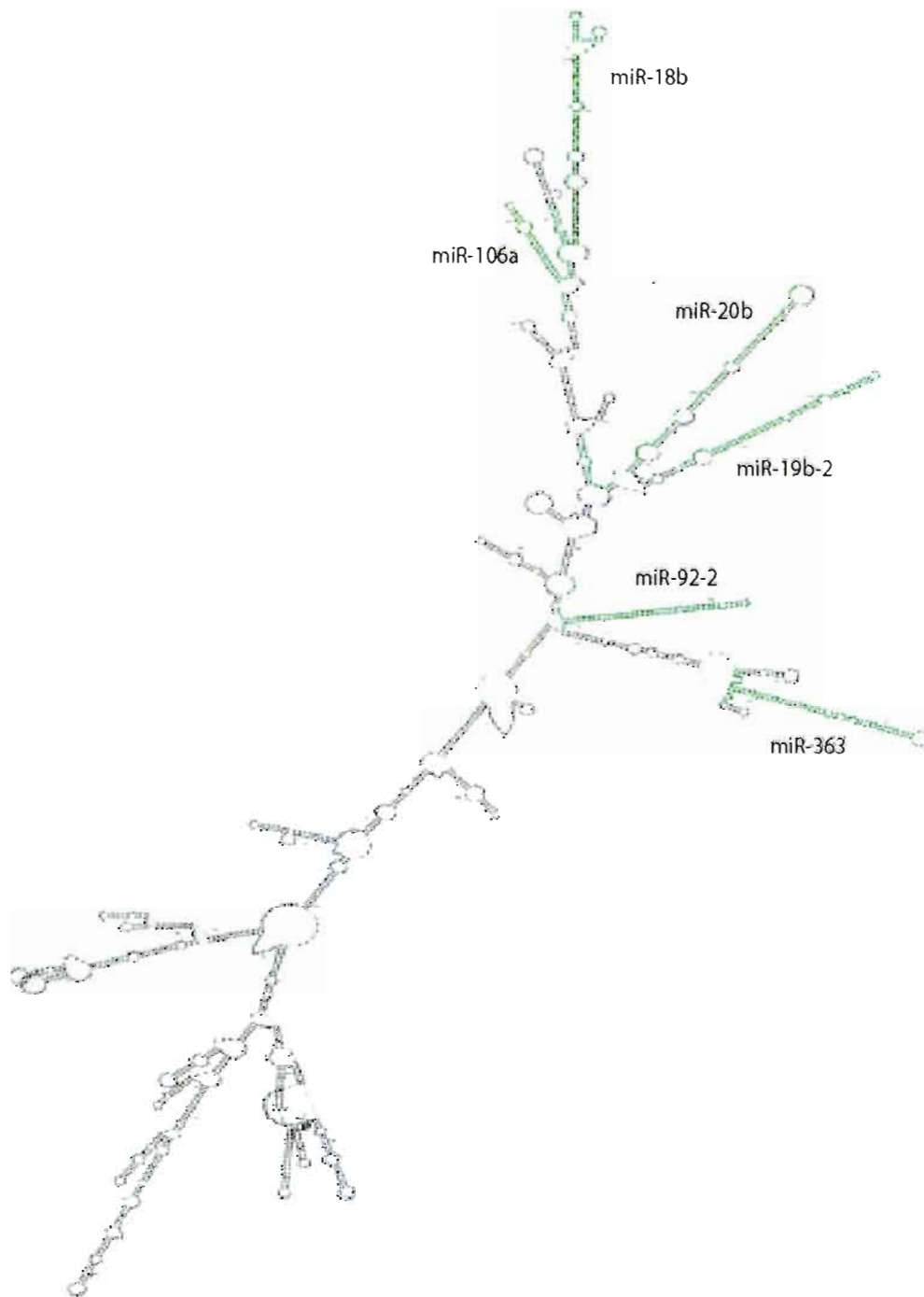




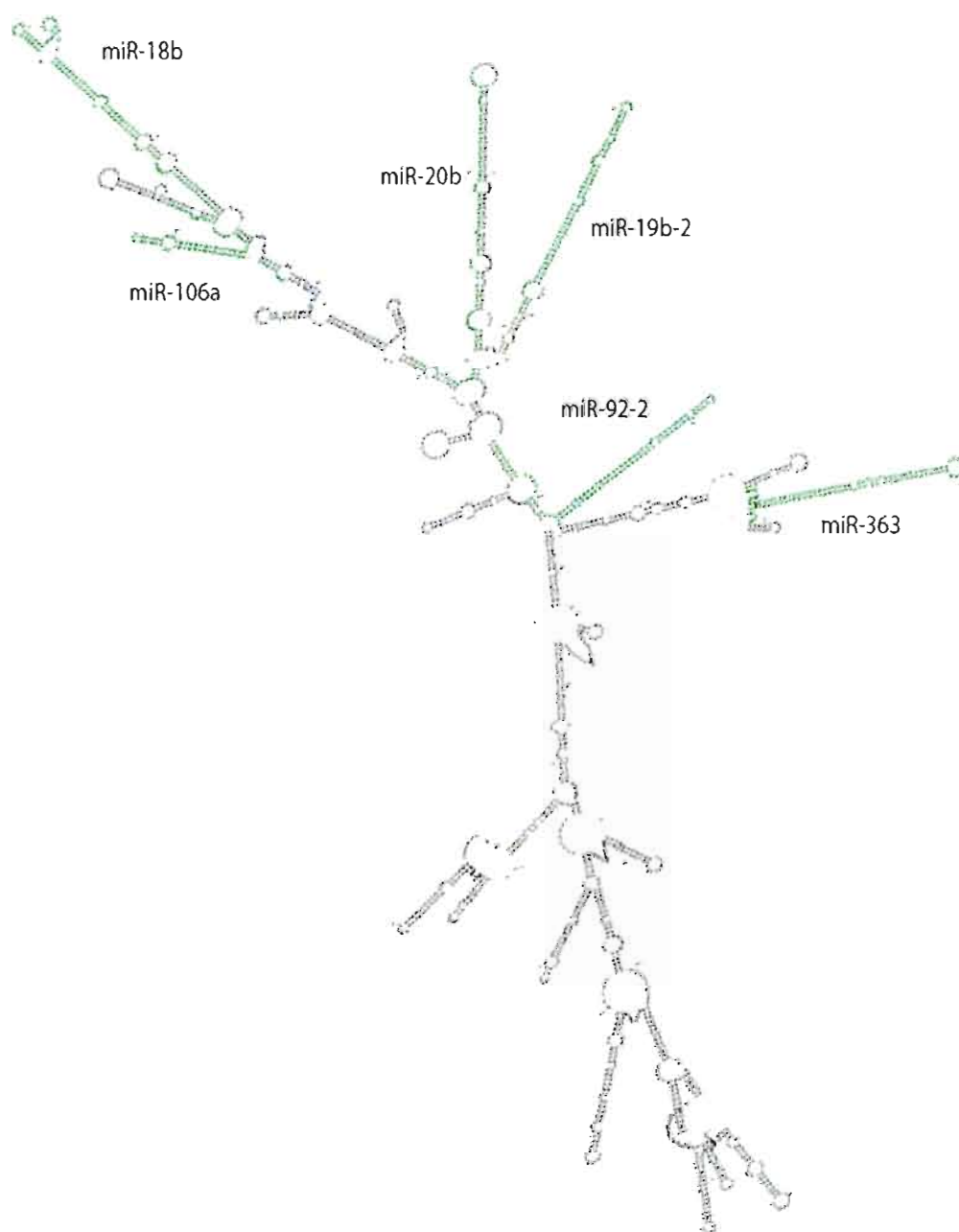
**Figure 4.8 Prédiction de la structure de l'ARN *Kis2 t2*, du pri-miARN complet correspondant au transcrit *Kis2 t2*. Les portions en vert correspondent aux tiges-boucles des miARN. Ici la tige-boucle de miR-18b n'est pas apparente, et notez celle de miR-20b qui ne correspond pas à la séquence connue (représenté par le mélange de deux couleurs). Seule la partie avec les miARN est représentée pour plus de clarté.**



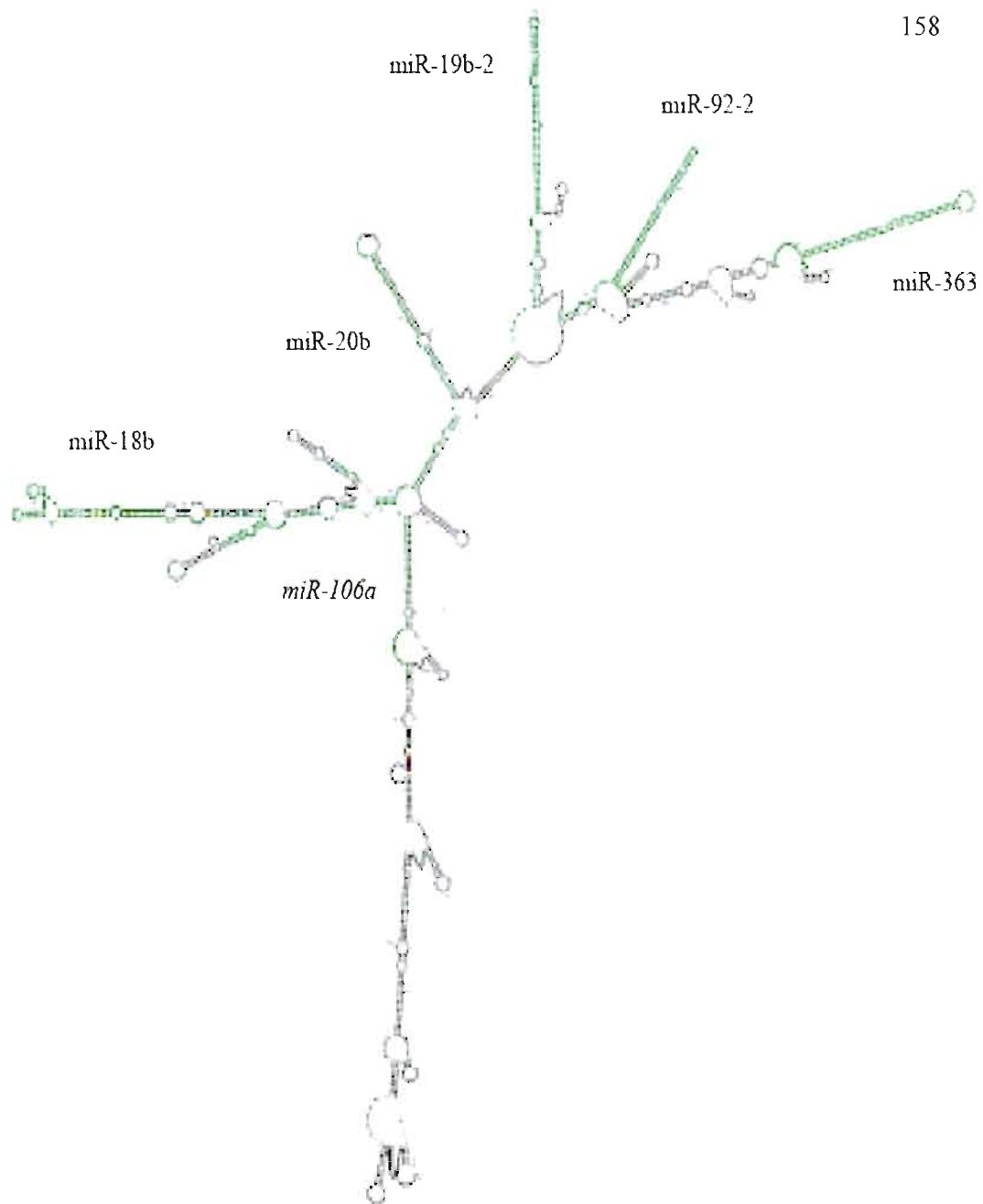
**Figure 4.9 Prédiction de la structure de l'ARN *Kis2* t3**, du pri-miARN complet correspondant au transcrit *Kis2* t3. Les portions en vert correspondent aux tiges-boucles des miARN. Ici la tige-boucle de miR-18b n'est pas apparente, et notez celle de miR-20b qui ne correspond pas à la séquence connue (représenté par le mélange de deux couleurs). Seule la partie avec les miARN est représentée pour plus de clarté.



**Figure 4.10** Prédiction de la structure de l'ARN *Kis2* t4, du pri-miARN complet correspondant au transcrit *Kis2* t4. Les portions en vert correspondent aux tiges-boucles des miARN. La tige-boucle de miR-18b est apparente, et notez celle de miR-20b qui ne correspond pas à la séquence connue (représenté par le mélange de deux couleurs). La structure complète est représentée.



**Figure 4.11 Prédiction de la structure de l'ARN *Kis2 t5*, du pri-miARN complet correspondant au transcrit *Kis2 t5*. Les portions en vert correspondent aux tiges-boucles des miARN. Ici la tige-boucle de miR-18b est apparente, et notez celle de miR-20b qui ne correspond pas à la séquence connue (représenté par le mélange de deux couleurs). La structure complète est représentée.**



**Figure 4.12 Prédiction de la structure de l'ARN *Kis2 t6*, du pri-miARN complet correspondant au transcrit *Kis2 t6*.** Les portions en vert correspondent aux tiges-boucles des miARN. Ici la tige-boucle de miR-106a disparaît, miR-18b est apparente, et notez celle de miR-20b qui ne correspond pas à la séquence connue (représenté par le mélange de deux couleurs). La structure complète est représentée.

#### 4.4 DISCUSSION

Dans le but de trouver la fonction des ARNnc *Kis2*, nous avons choisi d'identifier des protéines qui interagissent avec ces derniers. Nous avons trouvé les protéines hnRNP A2/B1, hnRNP A3, KSRP (et peut-être PSF) comme candidates potentielles. Les fonctions connues de ces protéines nous ont dès lors orienté vers la signification de la régulation post-transcriptionnelle des ARN *Kis2*, et notamment l'épissage. Nous proposons finalement un lien entre la régulation de l'épissage des pri-miARN *miR-106-363* et la régulation post-transcriptionnelle de l'expression des miARN de ce groupe.

Dans un premier temps, l'identification de ces protéines a été réalisée au moyen d'une expérience *in vitro*. La première question qui se pose face à ces résultats obtenus concerne donc leur spécificité. En effet, il y a toujours un risque que la présence forcée d'une grande quantité d'ARN d'une part, face à une grande quantité de protéines d'autre part, provoque des interactions plus artificielles que réelles. De plus, le contrôle négatif de cette expérience est loin d'être parfait. Ceci dit, parmi la multitude de protéines qui lient l'ARN dans une cellule, les protéines identifiées ont un lien avec la régulation de l'épissage. Étant donné la complexité du profil d'épissage des ARN *Kis2*, nous pensons donc qu'il s'agit plus que d'«heureuses» liaisons non spécifiques.

La spécificité de liaison des RBPs aux ARN est d'ailleurs l'objet de controverses. En effet, beaucoup de hnRNPs sont très abondantes dans la cellule (hnRNP A1 et C1 sont par exemple dix fois plus abondantes que le facteur général d'épissage U1, Dreyfuss 2002), et peuvent lier bon nombre de motifs nucléotidiques avec plus ou moins de spécificité. Au départ, il fut donc suggéré que ces protéines s'agglutinaient de façon non spécifique sur n'importe quel transcrit. À ce titre-là, les résultats concernant la liaison des hnRNPs aux ARN *Kis2* n'auraient pas grande valeur, sans compter que le profil de séparation de protéines obtenu sur gel semble être identique pour les cinq transcrits. Cette constatation fut au premier abord surprenante, car étant donné les différences qui existent entre ces cinq ARN, nous attendions un profil de liaison protéique distinct pour chacun d'eux. Cette similitude nous confrontait donc à un éventuel manque de spécificité de nos résultats. Ceci dit, les cinq transcrits ARN partagent des portions de séquences communes, qui peuvent être le siège d'interactions des protéines identifiées. Il peut paraître également surprenant que des protéines impliquées dans la régulation de

l'épissage soient capables de lier les transcrits épissés de la même manière que les non épissés. Là encore, on peut spéculer que les sites de liaison de ces protéines seraient plutôt exoniques qu'introniques. De plus, il semblerait que bon nombre de hnRNPs demeurent associées avec l'ARN une fois épissé (Daneholt 2001). Selon ces dernières considérations, il se pourrait donc que la présence équivalente de KSRP, hnRNP A2/B1 et hnRNP A3 pour les cinq transcrits *Kis2* ne soit pas de si mauvais augure.

Pour en revenir à la spécificité de liaison de ces protéines, il semblerait que leur apparente non-spécificité de liaison avec l'ARN ne soit qu'une façade derrière laquelle se cache au contraire un réseau d'interactions complexes et orchestrées. En fait, la nouvelle école de pensée stipule que chaque transcrit affiche un profil de RBPs qui lui est propre (Bennet 1992, Krecic 1999). Ainsi des protéines aussi communes que hnRNP A1, ou même PTB, peuvent effectivement s'associer avec de nombreux transcrits, et cependant conférer une spécificité en fonction des autres RBPs présentes (Dreyfuss 2002, Singh 2005). La spécificité d'action d'une RBP est donc difficilement appréciable dans l'individualité, et ne peut-être comprise que lorsqu'on considère l'ensemble de protéines dont elle fait partie. En plus du facteur présence/absence, la quantité de ces protéines est également importante. Par exemple, à faible quantité, les RBPs s'associeront préférentiellement avec un ARN de façon très spécifique sur une séquence nucléotidique conservée. Mais dû à leur flexibilité de liaison à l'ARN, ces mêmes RBPs à forte quantité auront tendance à également fixer des sites de liaisons moins spécifiques. Ces interactions pourraient être dénuées de sens. Or, ce n'est pas le cas, puisque leur présence peut interférer avec celle de RBPs liées plus spécifiquement, et ainsi bouleverser profondément la dynamique de régulation (Dreyfuss 2002, Singh 2005). Dans ces conditions, il apparaît donc difficile d'envisager le choix d'un quelconque ARN comme contrôle négatif de notre expérience. Finalement, la spécificité ou non spécificité de nos résultats apparaît plus que relative, et ne peut être envisagée sans le support d'expériences subséquentes.

À l'image de ces hypothèses, les ARN *Kis2* exhibent d'ailleurs un profil d'épissage différent dans les cellules NIH/3T3. Dans ce type cellulaire, tous les sites d'épissage sont activés, au contraire de ce qui est observé dans les tumeurs et les cellules BL/VL3. Ceci peut-être attribuable à un profil de RBPs différent entre les deux contextes cellulaires, que ce soit qualitativement ou quantitativement. Une première comparaison d'expression de ces protéines entre les cellules BL/VL3 et NIH/3T3 a révélé la présence de PTB, KSRP, hnRNP A2/B1 et hnRNP A3 dans ces deux types cellulaires avec

quelques différences notables. HnRNP A3 et PTB apparaissent surexprimées dans les cellules BL/VL3, alors que hnRNP A2/B1 apparaît sous-exprimée dans ces mêmes cellules. Si l'on considère qu'une modification dans le profil d'épissage peut résulter d'un déséquilibre dans les proportions de facteurs régulateurs, alors la différence d'épissage des ARN *Kis2* entre NIH/3T3 et BL/VL3 pourrait être corrélée avec une différence d'expression de hnRNP A2/B1, hnRNP A3 et PTB. Pour l'instant, ces observations ne constituent pas plus qu'un indice intéressant, et l'expérience mériterait d'être refaite de façon plus rigoureuse afin de réellement évaluer l'aspect quantitatif. De plus, cette différence dans le profil d'épissage peut être imputable à la présence/absence d'un facteur non identifié, qui pourrait collaborer avec nos protéines candidates. D'ailleurs, le fait que hnRNP A2/B1 soit clivée dans les extraits nucléaires des cellules NIH/3T3, et non dans ceux des BL/VL3, révèle que cette protéine est peut-être associée à des facteurs qui diffèrent entre ces deux types cellulaires. À ce titre, il serait intéressant de refaire l'expérience de purification des partenaires d'interaction des ARN *Kis2*, mais cette fois-ci sur une base comparative entre cellules BL/VL3 et NIH/3T3.

Afin de valider nos résultats, nous devons impérativement les confirmer par d'autres expériences, idéalement *in vivo*. Notamment, il faudrait trouver le moyen de purifier les RNPs relatives aux transcrits *Kis2* non épissés sans dégrader ces derniers. Les méthodes classiques d'extraction de protéines ont été essayées sans succès, mais il existe des kits d'extraction de RNPs que nous pourrions essayer. Quant à l'expérience de RIP elle-même, peut-être devrions-nous l'essayer en premier lieu sur des RNPs pour lesquelles l'expérience a déjà été paramétrée. Cela nous procurerait en même temps un contrôle positif. Dans un premier temps, j'ai réalisé l'expérience de RIP sans faire de cross-link selon des protocoles déjà établis (Mourelatos 2002). Il se peut que les RNPs que l'on cherche à purifier ne soient pas suffisamment stables, et nécessitent un cross link. Idéalement, avoir de grandes quantités d'anticorps spécifiques de ces protéines nous permettraient de tester plus de paramètres, notamment concernant le cross-link qui peut s'avérer difficile à paramétrer. De plus, cela nous permettrait aussi de conduire l'expérience à plus grande échelle (avec plus d'extrait cellulaire) et d'avoir plus de chance de détecter un signal. Alternativement, des expériences d'épissage *in vitro* des ARN *Kis2* impliquant hnRNP A2/B1, hnRNP A3, PTB et KSRP pourraient être envisagées, ou encore *in vivo* avec des siRNAs dirigés contre ces protéines.

En dehors de la régulation d'épissage, on pourrait penser à un rôle de ces



protéines dans le transport, puisque hnRNP A2/B1 et hnRNP A3 sont connues pour œuvrer dans le transport des ARN. C'est cependant peu probable car les ARN *Kis2* localisent dans le noyau. hnRNP A2/B1 est également connue pour affecter la stabilité des ARN, via les AREs. Alternativement cette protéine pourrait donc déstabiliser les ARN *Kis2*, et plusieurs séquences de types ARE (répétition de ATTTA) sont effectivement présentes dans ces ARN.

La régulation de l'épissage du gène *Kis2* nous amène directement à penser aux conséquences quant à la production de miARN du groupe *miR-106-363*. En effet, l'analyse des prédictions de structure secondaire des différents pri-miARN permet de mettre en évidence des différences de structures. Bien qu'il ne s'agisse que de prédictions, il n'en reste pas moins que les tiges-boucles relatives à chaque miARN sont clairement identifiables, ce qui renforce leur validité. Dans ces prédictions, la tige-boucle de miR-18b est absente des transcrits *Kis2* t1, t2 et t3, mais présente pour *Kis2* t4 et t5. Il est alors tentant de faire un parallèle entre la surexpression majoritaire de *Kis2* t1, t2 et t3 dans les tumeurs de souris, et l'absence de miR-18b. Par contre, cela ne justifie pas l'absence de miR-363 dont la tige-boucle est clairement prédite pour ces trois transcrits. Dans le cas de miR-363, on peut supposer qu'un autre moyen de régulation post-transcriptionnel est mis en jeu.

Un moyen simple de vérifier que les divers transcrits *Kis2* sont effectivement à l'origine d'une production de miARN différentes aurait été la transfection de chaque ADNc de ces transcrits dans des cellules NIH/3T3. L'analyse par Northern-blot du profil de miARN produits pour chacune des transfections nous aurait donné la réponse. Malheureusement, la production de miARN à partir des ADNc complets des pri-miARN de *miR-106-363*, sous la direction d'un promoteur CMV dans un vecteur d'expression de mammifère, n'a jamais fonctionné. L'expression des miARN à partir d'une construction ne contenant que la partie du pri-miARN avec les miARN ne fonctionne pas plus. En attendant de résoudre ce problème, l'hypothèse selon laquelle les différents transcrits *Kis2* sont à la base d'une régulation post-transcriptionnelle de la production des miARN nous tient toujours en haleine.

Si notre hypothèse devait être vérifiée, la régulation post-transcriptionnelle via l'épissage des pri-miARN *miR-106-363* serait liée à leur potentiel oncogénique. En effet une modification de la régulation de l'épissage résulterait donc en une production différentielle des miARN du groupe. Or, en fonction des miARN exprimés, les effets sur

la cellule ne seront pas nécessairement les mêmes. Par exemple, une combinaison de miARN du groupe *miR-106-363* pourrait s'avérer plus oncogénique que d'autres.

Pour finir, certaines des quatre protéines candidates ont d'ailleurs un lien avec le cancer. hnRNP A2/B1 est notamment impliquée dans plusieurs types de cancers, la plupart du temps par surexpression (Tableau 3.2, Carpenter 2005), et peut-être utilisée comme marqueur préventif lors des diagnostics du cancer du poumon (Tockman 1997, Qiao 1997). Le rôle de hnRNP A2/B1 dans le cancer n'est pas connu, et son potentiel oncogénique n'est pas vraiment prouvé. Toutefois, des siRNAs dirigés contre cette protéine provoquent l'apoptose des cellules transformées (Patry 2003), de même que la diminution de la prolifération d'une lignée cellulaire du colon (He 2005a). hnRNP A3 a également été rapportée comme étant surexprimée dans le cancer du colon (Ushigome 2005). Rappelons que miR-106a et miR-92 sont également surexprimés dans ce type de cancer (Volinia 2006).

L'effet de la surexpression des protéines hnRNP A2/B1 et A3 sur l'expression des miARN de ce groupe pourrait être étudié. Si ces protéines ont réellement un rôle dans la régulation post-transcriptionnelle des pri-miARN *miR-106-363*, alors un de leur lien avec le cancer pourrait être le groupe *miR-106-363*.

## DISCUSSION GÉNÉRALE ET PERSPECTIVES

L'objectif principal de ma thèse était la découverte de nouveaux oncogènes impliqués dans les leucémies de type T. Pour cela, la méthode d'étiquetage rétroviral a été appliquée, en utilisant le rétrovirus murin RadLV/VL3 qui induit des leucémies de type T avec une haute efficacité. L'application de cette méthode nous a permis de trouver un nouveau SCI rétroviral spécifique de ce type de leucémie, que l'on a baptisé *Kis2*. Nous avons mis en évidence que l'expression de deux gènes du locus *Kis2* était modifiée par les intégrations rétrovirales. L'un des deux gènes est un ARN non-codant, que nous avons caractérisé comme étant le pri-miARN du groupe de miARN *miR-106-363*. Sa surexpression dans les tumeurs de souris provoque l'accumulation de quatre miARN du groupe, et nous avons également démontré son implication dans des leucémies humaines. Nous démontrons un aspect du potentiel oncogénique de *miR-106-363* par sa capacité à induire l'indépendance d'ancrage dans des fibroblastes de souris. Et, finalement nous proposons trois gènes cibles potentiels de ces miARN, qui n'avaient jamais été impliqués auparavant dans les leucémies de type T, et dont l'expression dans les tumeurs de souris semble inversement corrélée à celle des miARN de *miR-106-363*.

### À propos des effets engendrés par les intégrations rétrovirales dans le génome hôte

Cette étude illustre avant tout la pertinence de la méthode d'étiquetage rétroviral. D'ailleurs, si jusqu'à présent elle n'avait permis de débusquer que des oncogènes relatifs aux cancers hématopoïétiques, cette technique a été nouvellement appliquée à la recherche de gènes impliqués dans des cancers du cerveau (Johansson 2004 et 2005). La puissance de l'étiquetage rétroviral s'est affirmée avec la publication du séquençage du génome de la souris en 2002, ce qui a permis un accès rapide au contenu génétique des SCIs et changé le cours de ma thèse. J'ai pu ainsi localiser le SCI *Kis2* dans le génome et

mettre en évidence non seulement la surexpression du gène *Kis2* situé tout proche, mais aussi celle du gène *Phf6* situé 300 Kpb en amont. Néanmoins, tous les gènes du locus *Kis2* ne sont pas affectés par les intégrations. Ce résultat démontre la complexité des effets engendrés par les insertions rétrovirales, qui peuvent se solder par la modification d'expression de plusieurs gènes à la fois, et ce sur de longues distances. D'ailleurs, le mécanisme d'activation à distance des rétrovirus n'est pas vraiment élucidé. Dans le cas de *Phf6* par exemple, il pourrait s'agir d'une activation par voie d'enhancer rétroviraux. Or rien n'a encore été prouvé quant à la portée dans le génome des séquences activatrices rétrovirales. Il a été proposé que l'intégration du rétrovirus à distance de gènes puisse agir sur leur expression en altérant des éléments de régulation importants, comme par exemple des éléments de contrôle de la conformation de la chromatine (Kim 2003).

Récemment, il a été rapporté que les intégrations rétrovirales dans le génome ne se faisaient pas totalement au hasard, et semblaient favoriser le 5' des gènes (Wu 2003). En conséquence, Wu *et al* émettent une mise en garde à propos de l'utilisation de rétrovirus murins dans les thérapies géniques, dont les intégrations préférentielles proches des gènes pourraient s'avérer cancéreuses. Ceci a d'ailleurs déjà été observé dans le cas d'enfants recevant un traitement par thérapie génique utilisant des rétrovirus murins: certains d'entre eux développaient des leucémies dues à l'insertion du rétrovirus en amont du gène *LMO2* (Check 2002, Kaiser 2003). L'étude du SCI *Kis2* montre que le problème n'est pas juste restreint au phénomène d'intégrations préférentiellement proches des gènes. Les intégrations rétrovirales peuvent engendrer des modifications plus complexes que la simple activation du gène adjacent, ce qui devrait être pris en compte lors de la considération des effets secondaires liés à l'utilisation des rétrovirus dans les thérapies géniques. Un des moyens utilisé pour investiguer cette complexité est de combiner l'étiquetage rétroviral avec des analyses de micropuces. L'expression de tous les gènes proches des sites intégrations peut alors être analysée à grande échelle, et des collaborations entre les gènes impliqués peuvent aussi être appréhendées. (Shin 2004, Johansson 2005).

### À propos du gène *Kis2/miR-106-363*

L'hypothèse de base concernant les SCIs étant qu'ils hébergent des gènes impliqués dans le cancer, les gènes *Kis2* et *Phf6* constituent donc deux nouveaux oncogènes potentiels. Je n'ai pas inclus *Phf6* dans la suite de mon étude, cependant, ce gène n'en reste pas moins un candidat intéressant pour de futures investigations. Je me suis consacrée à l'étude du gène *Kis2*, sur lequel rien n'était connu avant que son implication soit mise en évidence au niveau du SCI *Kis2*. Dans nos tumeurs, ce nouveau gène exhibait un profil d'épissage complexe, et sa caractéristique majeure était son caractère non-codant, c'est à dire l'impossibilité de décoder une quelconque séquence protéique qui ait du sens. Après avoir été inculquée durant toutes mes années universitaires par le dogme ADN-ARN-protéine, cette constatation fut au premier abord très intrigante. Nous pensions donc que l'information génétique dont nous disposions n'était pas complète : il devait y avoir une séquence codante quelque part. Après avoir complété l'information par RACE en 5' et en 3', la séquence restait cependant désespérément non-codante. La caractérisation du gène *Kis2* nous a donc initié au monde des ARN non-codants, et leur capacité à générer de l'information génétique sans traduction protéique. Les ARN non-codants ne sont pas bien (voire pas du tout) conservés entre espèce, et leurs fonctions dans l'état actuel de nos connaissances restent difficile à investiguer. Cela dit, l'explosion du domaine des miARN a éliminé une part de mystère, et nous a permis de mettre à jour que le gène *Kis2* n'était autre que le gène du groupe de miARN *miR-106-363*.

Pourquoi ne pas l'avoir découvert plus tôt? La principale cause réside en la probable grande efficacité de coupure du pri-miARN de *miR-106-363* par Drosha. Ceci ne permettait donc pas de détecter le pri-miARN complet en Northern-blot, en plus de l'amplification par 3' RACE erronée dont je n'explique toujours pas l'origine. Mes difficultés à identifier le gène *Kis2* comme le pri-miARN *miR-106-363* sont justifiées également par l'absence d'informations complètes quant à ce gène dans les banques de données. De plus, à l'époque, les informations de séquence de cette région n'étaient pas complètes.

Chez la souris, nous avons trouvé que le gène *Kis2* n'apparaît que faiblement exprimé dans la plupart des tissus, avec une plus forte expression dans le thymus et la

rate. Le groupe *miR-106-363* est apparenté au groupe *miR-17-92*, impliqué dans les leucémies de type B (He 2005), avec lequel il partage miR-19 et miR-92. Chez l'humain, nos résultats indiquent une expression constitutive de *miR-17-92* dans les lymphocytes circulants, ce qui n'est pas le cas de *miR-106-363*. Bien qu'ils contiennent des miARN en commun, leur profil d'expression n'est de toute évidence pas le même. La relation entre ces deux groupes reste donc à élucider. L'étude conjointe de leur expression dans différents tissus et différents types cellulaires, notamment au cours du développement, devrait être informative. Dans les leucémies humaines de type T, nous avons montré que le niveau d'expression de *miR-106-363* pouvait être fortement augmenté. Il serait intéressant de savoir quels sont les facteurs qui régissent cette augmentation. Par exemple, c-myc est-il capable de commander l'expression de *miR-106-363*, tout comme il le fait pour *miR-17-92* ? (O'Donnell 2005). Pour le savoir, des études sur le promoteur du gène *Kis2* pourraient être envisagées.

Pour l'instant, beaucoup de questions demeurent sans réponses en ce qui concerne les miARN. Un de leur mystère est la signification de leur redondance, complète ou partielle. Par exemple, *miR-17-92* et *miR-106-363* partagent non seulement miR-92 et miR-19, mais leurs autres miARN sont aussi très homologues au niveau de leur 5' seed. Selon les connaissances actuelles en matière de ciblage de gènes par les miARN, cela signifierait donc que toute cible prédite pour *miR-17-92* l'est également pour *miR-106-363*. On pourrait donc penser que tout effet oncogène relatif à *miR-17-92* peut être également rapporté à *miR-106-363*. En particulier, miR-106a et miR-20b du groupe *miR-106-363* peuvent-ils cibler les gènes E2F, de la même manière que miR-17-5p et miR-20a ? (O'Donnell 2005, Sylvestre 2006, Wood 2006) *miR-106-363* est-il également capable de favoriser l'angiogenèse, au même titre que *miR-17-92* ? (Dews 2006). Nous avons montré que miR-106a, miR-20b, miR-19 et miR-92 étaient capables d'induire l'indépendance d'ancrage dans des fibroblastes de souris, ce qui dénote une caractéristique de leur potentiel oncogène. Dans des analyses subséquentes de leur capacité oncogénique, les résultats rapportés pour *miR-17-92* pourront évidemment être investigués pour *miR-106-363*.

Nos résultats indiquent que la détection du pri-miARN n'était pas synonyme de la production de tous les miARN du groupe, mettant ainsi en évidence l'existence d'une régulation post-transcriptionnelle de leurs productions. Il apparaît donc clair qu'il ne suffit pas de détecter un pri-miARN pour déduire que ses miARN sont produits. C'est en quelque sorte le parallèle de ce que l'on observe avec les ARNm codant pour des

protéines. De la même manière, l'expression d'un miARN provenant d'un groupe n'implique pas nécessairement que les autres soient exprimés. Ceci indique donc que différentes combinaisons de miARN peuvent être produites à partir d'un même groupe. On peut imaginer que selon les combinaisons de miARN du groupe *miR-106-363* surexprimées, les conséquences sur la cellule ne seront pas les mêmes. Il serait intéressant de déchiffrer ces combinaisons sur un large échantillon de tumeurs et de leucémies, et d'analyser par la suite leur potentiel oncogénique. Certaines combinaisons pourraient être plus spécifiques d'un type de tumeur. Par exemple, miR-106a et miR-92 seulement sont surexprimés dans des tumeurs de la prostate et du colon (Volinia 2006). Ainsi, selon nos résultats dans les leucémies humaines, la surexpression de miR-19 apparaît plus souvent associée avec la surexpression de miR-92 qu'elle ne l'est avec celle de miR-106a.

Quel pourrait être le sens de ces combinaisons ? On pourrait penser simplement que des miARN sont exprimés conjointement pour cibler plus efficacement un même gène. Ceci peut être justifié par le fait que chez les mammifères, les 3' UTRs des gènes contiennent rarement plus d'un site pour un miARN donné, ce qui est censé être pauvrement efficace (Hornstein 2006, Stark 2005). Par contre, ils contiennent souvent une combinaison de sites uniques pour différents miARN. Dans notre étude, nous avons essayé à plusieurs reprises de montrer une éventuelle synergie entre miR-106a, miR-20b, miR-19 et miR-92. Nous n'avons pas obtenu de résultats allant dans le sens d'une synergie, ce qui nous a surpris. Cependant, cette hypothèse mériterait d'être investiguée plus profondément. Les expériences pourraient être refaites, mais en essayant une plus grande gamme de concentration de miARN que nous ne l'avons fait. D'autre part, il est possible que les concentrations de chaque miARN aient une importance. Nous observons d'ailleurs, chez la souris et l'humain, que les miARN de *miR-106-363* ne sont pas exprimés dans les mêmes proportions. Cela vaudrait la peine de refaire les expériences en variant les concentrations de chaque miARN individuellement. Également, il serait intéressant de tester l'effet de différentes combinaisons, comme par exemple miR-106a avec miR-20b, ou miR-19 avec miR-92, *etc.*

L'effet des miARN dans une cellule passe par l'inhibition de la traduction de leurs gènes cibles. Dans le cas d'une surexpression de miARN dans un contexte tumoral, la logique veut donc que les cibles impliquées soient de type suppresseur de tumeurs. Parmi les prédictions de cibles disponibles pour le groupe *miR-106-363*, certaines sont effectivement des suppresseurs de tumeurs. Quelques-unes sont parfois célèbres, comme le gène Rb par exemple, qui a d'ailleurs été impliqué en tant que cible de miR-106a

(Volinia 2006). Ceci n'apparaît pas si évident dans nos tumeurs de souris, mais il faut dire aussi que Rb fait l'objet de nombreux types de régulations post-transcriptionnelles. Le ciblage de Rb par miR-106a mériterait donc de plus amples études.

Les cibles prédites individuellement pour miR-106a/20b, miR-19b, et miR-92 sont nombreuses (672, 456 et 357, respectivement). Dans mon étude, j'ai décidé de ne considérer que les cibles communes aux quatre miARN, pour me limiter aux prédictions les plus significatives à l'échelle du groupe *miR-106-363*. Nous avons montré que les gènes *Rbp1-like*, *Myliip* et probablement *Hipk3* pouvaient être ciblés par les miARN de *miR-106-363*. Si *Rbp1-like* est connu pour avoir des caractéristiques de suppresseur de tumeur, *Myliip* et *Hipk3* ne sont pas encore connus sous cet angle. Quoiqu'il en soit, nous sommes les premiers à impliquer ces trois gènes dans un processus tumoral. À mon sens, cette hypothèse mériterait des investigations expérimentales prioritaires pour être prouvée. Par exemple, on pourrait vérifier si le ciblage de ces gènes est à l'origine du phénotype d'indépendance d'ancrage induit par *miR-106-363*. Une expérience d'indépendance d'ancrage mettant en scène des siARN spécifiques de chacun de ces gènes devrait être conduite, en particulier pour *Myliip* qui est connu pour interagir avec des protéines du cytosquelette (Olson 1999).

Alternativement, l'implication des miARN dans la tumorigenèse est sans doute multiple. D'autres gènes cibles ayant des caractéristiques de suppresseurs de tumeurs figurent parmi les prédictions, comme par exemple *RBL2*, cible exclusive des 5' seed du type de miR-106a. Il n'est cependant pas possible de tester toutes les prédictions individuellement. D'ailleurs, il est peut-être fortuit de considérer l'effet oncogénique d'un groupe de miARN à l'échelle d'un ou deux gènes seulement. Nous savons aussi que si certains oncogènes majeurs agissent avec fracas, d'autres exercent leurs effets plus subtilement, notamment par le biais de collaborations. Les approches de choix pour appréhender l'effet des "oncomirs" seront donc celles qui considéreront l'ensemble du protéome.

Finalement, on peut se demander quel est le rôle normal de *miR-106-363*. Si l'on porte attention à la nature des gènes cibles de *miR-106-363*, il n'y a pas de fonction générale qui ressort de l'ensemble. On peut toutefois essayer de raisonner sur le mode de l'exclusion mutuelle (voir Chapitre I, 1.2.2.2). Selon ce modèle, le gène *Kis2* est plus fortement exprimé dans le thymus chez la souris, et cette expression viserait à en préserver l'identité tissulaire. Si on analyse les quarante premières cibles prédites individuellement pour chaque miARN, 15% sont reliées au cerveau pour miR-106a et



miR-92, et 22% pour miR-19. Pour les prédictions communes à ces trois miARN, le chiffre augmente à 30%. La fonction cérébrale est donc celle qui revient le plus souvent dans les meilleurs scores de prédictions si l'on compare avec d'autres types de fonctions. Toutefois, cela ne m'apparaît pas assez flagrant pour associer *miR-106-363* au modèle d'exclusion mutuelle, comme dans le cas de miR-1 et miR-124 par exemple (Hornstein 2006). Chez l'humain, nous avons vu que *miR-106-363* n'est pas détectable dans le thymus adulte, et l'est très faiblement dans les PBMCs. Peut-être est-il plus fortement exprimé à certains moments du développement. Parmi les cibles de ces miARN, on retrouve souvent des gènes plutôt impliqués dans la prolifération et le cycle cellulaire, comme Rb par exemple. *MiR-106-363* servirait-il de «garde fou» quant à l'expression de ces gènes ?

D'ailleurs, nous avons constaté que Phf6 figurait parmi les prédictions de miR-106a. Or, nous avons vu que les intégrations rétrovirales au niveau du locus *Kis2* activent conjointement l'expression de *miR-106-363* et Phf6. Ceci peut donc signifier qu'une activation «normale» de la chromatine à cet endroit favorise la transcription de ces deux gènes. On peut dès lors imaginer que l'expression de miR-106a soit couplée à celle de Phf6, pour prévenir les débordements d'expression de ce dernier comme le fait *miR-17-92* avec les facteurs E2Fs (Sylvestre 2006, Wood 2006). Évidemment, ce ciblage de Phf6 par miR-106a reste à prouver, et pourra être considéré lorsque le gène Phf6 sera mieux caractérisé.

### À propos du sens de l'épissage alternatif du gène *Kis2*

Le dernier volet de cette thèse concerne l'identification des RBPs qui interagissent les ARN *Kis2*. Dans un premier temps, comme nous ignorions l'existence de *miR-106-363*, ce travail avait pour but de nous renseigner sur la fonction des ARN *Kis2*. Par la suite, la nature des RBPs identifiées nous a orienté vers la régulation de l'épissage alternatif des pri-miARN de *miR-106-363*. Mon hypothèse de fond combine le profil d'épissage complexe du gène *Kis2* avec l'expression différentielle des miARN du groupe *miR-106-363*. L'épissage alternatif constituerait une régulation post-transcriptionnelle à l'origine des différentes combinaisons de miARN produits, comme il est à la base de la diversité protéique engendrée par un seul gène. Les RBPs identifiées (hnRNP A2/B1,

hnRNP A3, KSRP et PTB) sont d'excellentes candidates en tant que médiatrices de cette régulation.

De nombreux points restent à prouver, notamment que ces protéines interagissent bien avec les pri-miARN de *miR-106-363 in vivo*. Un des problèmes à résoudre de façon imminente est celui de l'expression dans les cellules eucaryotes des miARN à partir d'une construction contenant le pri-miARN *miR-106-363* au complet. Ceci constitue d'ailleurs le projet d'une nouvelle étudiante dans notre laboratoire, et s'avère primordiale pour investiguer l'hypothèse du rôle de l'épissage alternatif dans la régulation post-transcriptionnelle du groupe *miR-106-363*.



## RÉFÉRENCES BIBLIOGRAPHIQUES

- Aasland, R., T.J. Gibson, et A.F Stewart. 1995. "The PHD finger : implication for chromatin-mediated transcriptional regulation". *Trends in Biochemical Science*, Vol. 20, p. 56-59.
- Akagi, K., T. Suzuki, R. M. Stephens, N. A. Jenkins et N. G. Copeland. 2004. "RTCGD: retroviral tagged cancer gene database". *Nucleic Acids Research*, Vol. 32, D523-D527.
- Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter. 2002. "Molecular Biology of the cell". *Garlan science*, New York and London, 4th edition.
- Alvarez-Garcia, I., et A. E. Miska. 2005. "MicroRNA functions in animal development and human disease". *Development*, Vol. 132, p. 4653-4662.
- Ambros, V., B. Bartel, D. P. Bartel, C. B. Burge, J. C. Carrington, X. Chen , G. Dreyfuss , S. R. Eddy, S. Griffiths-Jones, M. Marshall, M. Matzke, G. Ruvkun, et T. Tuschl. 2003. "A uniform system for microRNA annotation". *RNA*, Vol. 9, n°3, p. 277-279.
- Ambros, V. 2004. "The functions of animal microRNAs". *Nature*, Vol. 431, p. 350- 355.
- Aravin, A. A., M. Lagos-Quintana, A. Yalcin, M. Zavolan, D. Marks, B. Snyder, T. Gaasterland, J. Meyer, et T. Tuschl. 2003. "The small RNA profile during *Drosophila melanogaster* development". *Developmental Cell*, Vol. 5, p. 337-350.
- Bagga, S., J. Bracht, S. Hunter, K. Massirer, J. Holtz, R. Eachus, et A. E. Pasquinelli. 2005. "Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation". *Cell*, Vol. 122, n°4, p. 553-563.
- Bartel, D. P. 2004. "MicroRNAs: genomics, biogenesis, mechanism, and function". *Cell*, Vol. 116, n°2, p. 281-297.
- Bashirullah, A., A. E Pasquinelli, A. A. Kiger, N. Perrimon, G. Ruvkun, et C. S. Thummel. 2003. "Coordinate regulation of small temporal RNAs at the onset of *Drosophila* metamorphosis". *Developmental Biology*, Vol. 259, p. 1-8.
- Ben-David, Y., E. Yefenof, et M. Kotler. 1987. "Clonal analysis of radiation leukemia virus-induced leukemic and preleukemic murine cells". *Cancer Research*, Vol. 47, p. 6590-6594.
- Bennett, M., S. Piñol-Roma, D. Staknis, G. Dreyfuss, et R. Reed. 1992. «Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. *Molecular and Cellular Biology*, Vol. 12, n°7, p. 3165-3175.

- Berget, S. M. 1995. "Exon recognition in vertebrate splicing". *Journal of Biological Chemistry*, Vol. 270, n°6, p. 2411-2414.
- Bernstein, E., A.A. Caudy, S.M. Hammond et G.J. Hannon. 2001. "Role for a bidentate ribonuclease in the initiation step of RNA interference." *Nature*, Vol. 409, p. 363-366.
- Bernstein, E., S. Y. Kim, M. A. Carmell, E. P. Murchison, H. Alcorn, M. Z. Li, A. A. Mills, S. J. Elledge, K. V. Anderson, et G. J. Hannon. 2003. "Dicer is essential for mouse development". *Nature Genetics*, Vol. 35, n°3, p. 215-217.
- Bilodeau, P. S., J. K. Domsic, A. Mayeda, A. R. Krainer, et C. M. Stoltzfus. 2001. «RNA splicing at human immunodeficiency virus type 1 3' splice site A2 is regulated by binding of hnRNP A/B proteins to an exonic splicing silencer element». *Journal of Virology*, Vol. 75, p. 8487-8497.
- Binda, O., J.S. Roy, et P.E. Branton. 2006. "RBP1 family proteins exhibit SUMOylation-dependent transcriptional repression and induce cell growth inhibition reminiscent of senescence." *Molecular and Cellular Biology*, Vol. 26, p. 1917-1931.
- Birren, B., E. D. Green, S. Klaplitz, R. M. Myers, et J. Roskams. 1998. Chapitre 2 dans "Genome analysis: a laboratory manual, volume 2". Cold Spring Harbor Laboratory press, p. 80-82.
- Björns P., M. Halonen, J.J. Palvimo, M. Kolmer, J. Aaltonen, P. Ellonen, J. Perheentupa, I. Ulmanen, et L. Peltonen. 2000. "Mutations in the AIRE gene: effects on subcellular location and transactivation function of the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy protein". *American Journal of Human Genetic*, Vol. 66, p. 378-392.
- Black, D. L. 1992. "Activation of c-src neuron-specific splicing by an unusual RNA element in vivo and in vitro". *Cell*, Vol. 69, n°5, p. 795-807.
- Black, D. L. 2000. "Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology". *Cell*, Vol. 103, p. 367-370.
- Black, D. L. 2003. "Mechanisms of alternative pre-messenger RNA splicing". *Annual review of biochemistry*, Vol. 72, p. 291-336.
- Borchert, G. M., W. Lanier, et B. L. Davidson. 2006. "RNA polymerase III transcribes human microRNAs". *Nature structural and molecular biology*, Vol. 13, n°12, p. 1097-1101.
- Bordoli, L., S. Husser, U. Luthi, M. Netsch, H. Osmani, et R. Eckner. 1992. "Functional analysis of the p300 acetyltransferase domain : the PHD finger of p300 but not of CBP is dispensable for enzymatic activity". *Nucleic Acids Research*, Vol. 29, p. 4462-4471.
- Börjeson, M., H. Forssman, et O. Lehmann. 1962. "An X-linked, recessively inherited syndrome characterized by grave mental deficiency, epilepsy, and endocrine disorder". *Acta in Medicin Scandinavy*, Vol. 171, p. 13-21.

- Brennecke, J., A. Stark, R. B. Russell, et S. M. Cohen. 2005. "Principles of microRNA-target recognition". *PLoS Biology*, Vol. 3, n°3, e85.
- Bretscher, A., D. Reczek, et M. Berryman. 1997. "Ezrin: a protein requiring conformational activation to link microfilaments to the plasma membrane in the assembly of cell surface structures." *Journal of Cell Science*, Vol. 110, p. 3011-3018.
- Brockdorff, N., A. Ashworth, G. F. Kay, V. M. McCabe, D. P. Norris, P. J. Cooper, S. Swift, et S. Rastan. 1992. "The product of the mouse Xist gene is a 15 Kb inactive X-specific transcript containing no conserved ORF and located in the nucleus". *Cell*, Vol. 71, n°3, p. 515-526.
- Burd, C. G., M. S. Swanson, M. Gorlach, et G. Dreyfuss. 1989. "Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: a diversity of RNA binding proteins is generated by small peptide inserts". *Proceedings of National Academy of Science USA*, Vol. 86, p. 9788-9792.
- Burge, C. B., T. Tuschl, et P. A. Sharp. 1999. "The RNA World". *Édition Gesteland, R. F., T. R. Cech, et J. F. Atkins, Cold Spring Harbor* (2<sup>nd</sup> édition), p. 525-560.
- Caceres, J. F., S. Stamm, D. M. Helfman, et A. R. Krainer. 1994. "Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors". *Science*, Vol. 265, p. 1706-1709.
- Cai, X., C.H. Hagedorn, et B.R. Cullen. 2004. "Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs." *RNA*, Vol. 10, p. 1957-1966.
- Calin, G. A., C. Sevignani, C. D. Dumitru, T. Hyslop, E. Noch, S. Yendamuri, M. Shimizu, S. Rattan, F. Bullrich, M. Negrini, et C. M. Croce. 2004. "Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers". *Proceedings of National Academy of Science U. S. A.*, Vol. 101, p. 2999-3004 a.
- Calin, G. A., C. G. Liu, C. Sevignani, M. Ferracin, N. Felli, C. D. Dumitru, M. Shimizu, A. Cimmino, S. Zupo, M. Dono, M. L. Dell'Aquila, H. Alder, L. Rassenti, T. J. Kipps, F. Bullrich, M. Negrini, et C. M. Croce. 2004. "MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias". *Proceedings of National Academy of Science US A*, Vol. 101, n°32, p. 11755-11760 b.
- Calin, G. A., M. Ferracin, A. Cimmino, G. Di Leva, M. Shimizu, S. E. Wojcik, M. V. Iorio, R. Visone, N. I. Sever, M. Fabbri, R. Iuliano, T. Palumbo, F. Pichiorri, C. Roldo, R. Garzon, C. Sevignani, L. Rassenti, H. Alder, S. Volinia, C. G. Liu, T. J. Kipps, M. Negrini, et C. M. Croce. 2005. "A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia". *New England Journal of Medicine*, Vol. 353, n°17, p. 1793-1801.
- Calin, G. A., et C. M. Croce. 2006. "MicroRNA signatures in human cancers". *Nature Reviews Cancer*, Vol. 6, p. 857-86.

- Caputi, M., A. Mayeda, A. R. Krainer, et A. M. Zahler. 1999. "hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing". *EMBO Journal*, Vol. 18, p. 4060–4067.
- Carmo-Foncesca, M., L. Mendes-soares et I. Campos. 2000. "To be or not to be in the nucleolus". *Nature and Cell Biology*, Vol. 2, p. 107-112.
- Carpenter, B., C. MacKay, A. Alnabulsi, M. MacKay, C. Telfer, W. T. Melvin, et G. I. Murray. 2006. "The roles of heterogeneous nuclear ribonucleoproteins in tumour development and progression". *Biochimica et Biophysica Acta*, Vol. 1765, p. 85–100.
- Carter, D., L. Chakalova, C. S. Osborne, Y. Dai, et P. Fraser. 2002. "Long-range chromatin regulatory interactions in vivo". *Nature Genetics*, Vol. 32, p. 623-626.
- Cavanagh, M.H., S. Landry, B. Audet, C. Arpin-Andre, P. Hivin, M.E. Pare, J. Thete, E. Wattel, S.J. Marriott, J.M. Mesnard, et B. Barbeau. 2006. "HTLV-I antisense transcripts initiating in the 3'LTR are alternatively spliced and polyadenylated." *Retrovirology*, Vol. 3, p. 15.
- Cawley, S., S. Bekiranov, H. H. Ng, P. Kapranov, E. A. Sekinger, D. Kampa, A. Piccolboni, V. Sementchenko, J. Cheng, A. J. William, R. Wheeler, B. Wong, J. Drenkow, M. Yamanaka, S. Patel, S. Brubaker, H. Tammana, G. Helt, K. Struhl, et T. R. Gingeras. 2004. "Unbiased Mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs". *Cell*, Vol. 116, p. 499-509.
- Check, E. 2002. "A tragic setback". *Nature*, Vol. 420, p. 116-118.
- Chesi, M., E. Nardini, R.S. Lim, K.D. Smith, W.M. Kuehl, et P.L. Bergsagel. 1998. "The t(4;14) translocation in myeloma dysregulates both FGFR3 and a novel gene, MMSET, resulting in IgH/MMSET hybrid transcripts". *Blood*, Vol. 92, p.3025-3034.
- Chou, C. F., A. Mulky, S. Maitra, W. J. Lin, R. Gherzi R, J. Kappes, et C. Y. Chen. 2006. "Tethering KSRP, a decay-promoting AU-rich element-binding protein, to mRNAs elicits mRNA decay". *Molecular and Cellular Biology*, Vol. 26, n°10, p. 3695-3706.
- Chu, C. Y., et T. M. Rana. 2006. "Translation Repression in Human Cells by MicroRNA-Induced Gene Silencing Requires RCK/p54". *PLoS Biology*, Vol. 4, n°7, e210.
- Cimmino, A., G. A. Calin, M. Fabbri, M. V. Iorio, M. Ferracin, M. Shimizu, S. E. Wojcik, R. I. Aqeilan, S. Zupo, M. Dono, L. Rassenti, H. Alder, S. Volinia, C. G. Liu, T. J. Kipps, M. Negrini, et C. M. Croce. 2005. "miR-15 and miR-16 induce apoptosis by targeting BCL2". *Proceedings of National Academy of Science U S A*, Vol. 102, n°39, p. 13944-13949.
- Clurman, B. E., et W. S. Hayward. 1989. "Multiple proto-oncogene activations in avian leukemia virus-induced lymphomas: evidence for stage-specific events". *Molecular and Cellular Biology*, Vol. 9, n°6, p. 2657-2664.

- Coffin, J. M., S. H. Hughes, et H. E. Varmus. 1997. "Retroviruses". *Plainview (NY): Cold Spring Harbor Laboratory Press*.
- Corcoran, L.M., J.M. Adams, A.R. Dunn, et S. Cory. 1984. "Murine T lymphomas in which the cellular myc oncogene has been activated by retroviral insertion". *Cell*, Vol. 37, p. 113-122.
- Costinean, S., N. Zanesi, Y. Pekarsky, E. Tili, S. Volinia, N. Heerema, et C. M. Croce. 2006. "Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice". *Proceedings of National Academy of Science U S A*, Vol. 103, n°18, p. 7024-7029.
- Cui, D., G. Jin, T. Gao, T. Sun, F. Tian, G.G. Estrada, H. Gao, A. Sarai. 2004. "Characterization of BRCA1 and its novel antigen epitope identification." *Cancer Epidemiology Biomarkers Preview*, Vol. 13, p. 1136-1145.
- Cuyper, H. T., G. Selten, W. Quint, M. Zijlstra, E. R. Maandag, W. Boelens, P. V. Wezenbeek, C. Melief, et A. Berns. 1984. "Murine leukemia virus-induced T-cell lymphomagenesis : integration of proviruses in a distinct chromosomal region". *Cell*, Vol. 37, p. 141-150.
- Daneholt, B. 2001. "Assembly and transport of a premessenger RNP particle". *Proceedings of National Academy of Science USA*, Vol. 98, n°13, p. 7012-7017.
- Decleve, A., M. Lieberman, J. N. Ihle, P. N. Rosenthal, M. L. Lung, et H. S. Kaplan. 1978. "Physicochemical, biological and serological properties of a leukemogenic virus isolated from cultured RadLV-induced lymphomas of C57BL/Ka mice". *Virology*, Vol. 90, n°1, p. 23-35.
- Denicourt, C., C.A. Kozak, et E. Rassart. 2003. "Gris1, a new common integration site in Graffi murine leukemia virus-induced leukemias: overexpression of a truncated cyclin D2 due to alternative splicing". *Journal of Virology*, Vol. 77, p. 37-44.
- Dews, M., A. Homayouni, D. Yu, D. Murphy, C. Seignani, E. Wentzel, E. E. Furth, W. M. Lee, G. H. Enders, J. T. Mendell, et A. Thomas-Tikhonenko. 2006. "Augmentation of tumor angiogenesis by a Myc-activated microRNA cluster". *Nature Genetics*, Vol. 38, n°9, p. 1060-1065.
- Dignam, J. D., R. M. Lebovitz, et R. G. Roeder. 1983. "Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei". *Nucleic Acids Research*, Vol. 11, p. 1475-1489.
- Doench, J. G., et P. A. Sharp. 2004. "Specificity of microRNA target selection in translational repression". *Genes and development*, Vol. 18, p. 504-511.
- Dreyfuss, G., V. N. Kim, et N. Kataoka. 2002. "Messenger-RNA-binding proteins and the messages they carry". *Nature Reviews Molecular Cell Biology*, Vol. 3, p. 195-205.



- Dudley, J.P., J.A Mertz, , M. Lozano, et Dr Broussard. 2002. "What retroviruses teach us about the involvement of c-Myc in leukemias and lymphomas." *Leukemia*, Vol. 16, p. 1086-1098.
- Eis, P. S., W. Tam, L. Sun, A. Chadburn, Z. Li, M. F. Gomez, E. Lund, et J. E. Dahlberg. 2005. "Accumulation of miR-155 and BIC RNA in human B cell lymphomas". *Proceedings of National Academy of Science U S A*, Vol. 102, n°10, p. 3627-3632.
- Ellerman, V., et O. Bang. 1908. "Experimentelle Leukämie bei Hühnern". *Zentralbl Bakteriol Parasitenkd Infektionskr Hyg Abt Orig*, Vol. 46, p. 595.
- Faller, M., M. Matsunaga, S. Yin, J. A. Loo, et F. Guo. 2006. "Heme is involved in microRNA processing". *Nature structural and molecular biology*, advanced online publication.
- Farh, K. K., A. Grimson, C. Jan, B. P. Lewis, W. K. Johnston, L. P. Lim, C. B. Burge, et D. P. Bartel. 2005. "The widespread impact of mammalian microRNAs on mRNA repression and evolution". *Science*, Vol. 310, p. 1817-1821.
- Felli, N., L. Fontana, E. Pelosi, R. Botta, D. Bonci, F. Facchiano, F. Liuzzi, V. Lulli, O. Morsilli, S. Santoro, M. Valtieri, G. A. Calin, C. G. Liu, A. Sorrentino, C. M. Croce, et C. Peschle. 2005. "MicroRNAs 221 and 222 inhibit normal erythropoiesis and erythroleukemic cell growth via kit receptor down-modulation". *Proceedings of National Academy of Science U S A*, Vol. 102, n°50, p. 18081-18086.
- Filmus, J., et S.B. Selleck. 2001. "Glypicans : proteoglycans with a surprise". *Journal of Clinical Investigation*, Vol. 108, p. 497-501.
- Fleischer, T.C., U.J. Yun, et D.E. Ayer. 2003. "Identification and characterization of three new components of the mSin3A corepressor complex." *Molecular and Cellular Biology*, Vol. 23, p. 3456-3467.
- Friend, C. 1957. "Cell-free transmission in adult Swiss mice of a disease having the character of a leukemia". *Journal of Experimental Medicine*, Vol. 105, p. 307-318.
- Friend, C. 1957. "Leukemia of adult mice caused by a transmissible agent". *Ann N Y Acad Sci*. Vol. 68, p. 522-532.
- Garzon, R., M. Fabbri, A. Cimmino, G. A. Calin, et C. M. Croce. 2006. "MicroRNA expression and function in cancer". *Trends in Molecular Medecin*, Vol. 12, p. 580-587.
- Gherzi, R., K. Y. Lee, P. Briata, D. Wegmuller, C. Moroni, M. Karin, et C. Y. Chen. 2004. "A KH domain RNA binding protein, KSRP, promotes ARE-directed mRNA turnover by recruiting the degradation machinery". *Molecular Cell*, Vol. 14, n°5, p. 571-583.
- Gibbons, R.J., S. Bachoo, D.J. Picketts, S. Aftimos, B. Asenbauer, J. Bergoffen, S.A. Berry, N. Dahl, A. Fryer, K. Keppler, K. Kurosawa, M.L. Levin, M. Masuno, G. Neri, M.E. Pierpont, S.F. Slaney, et D.R. Higgs. 1997. "Mutations in transcriptional

- Gibbons, R.J., S. Bachoo, D.J. Picketts, S. Aftimos, B. Asenbauer, J. Bergoffen, S.A. Berry, N. Dahl, A. Fryer, K. Keppler, K. Kurosawa, M.L. Levin, M. Masuno, G. Neri, M.E. Pierpont, S.F. Slaney, et D.R. Higgs. 1997. "Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain." *Nature Genetics*, Vol. 17, p. 146-148.
- Giraldez, A. J., R. M. Cinalli, M. E. Glasner, A. J. Enright, J. M. Thomson, S. Baskerville, S. M. Hammond, D. P. Bartel, et A. F. Schier. 2005. "MicroRNAs regulate brain morphogenesis in zebrafish". *Science*, Vol. 308, p. 833-838.
- Girard, L., Z. Hanna, N. Beaulieu, C.D. Hoemann, C. Simard, C.A. Kozak, et P. Jolicoeur. 1996. "Frequent provirus insertional mutagenesis of Notch1 in thymomas of MMTVD/myc transgenic mice suggests a collaboration of c-myc and Notch1 for oncogenesis." *Genes and Development*, Vol. 10, p. 1930-1944.
- Gozani, O., J. G. Patton, and R. Reed. 1994. "A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction". *EMBO Journal*, Vol. 13, p. 3356-3367.
- Gregory, R. I., P. Thimmaiah, N. C. Chendrimada, et R. Shiekhattar. 2005. "Human RISC Couples MicroRNA Biogenesis and Posttranscriptional Gene Silencing". *Cell*, Vol. 123, issue 4, p. 631-640.
- Gribnau, J., K. Diderich, S. Pruzina, R. Calzolari, et P. Fraser. 2000. "Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus." *Molecular Cell*, Vol. 5, p. 377-386.
- Griffin, M. E., B. J. Hamilton, K. M. Roy, M. Du, A. M. Willson, B. J. Keenan, X. W. Wang, et R. C. Nichols. 2004. "Post-transcriptional regulation of glucose transporter-1 by an AU-rich element in the 3'UTR and by hnRNP A2". *Biochemical and Biophysical Research Communications*, Vol. 318, p. 977-982.
- Griffiths-Jones, S. 2004. "The microRNAs registry". *Nucleic Acids Research*, Vol. 3, D109-D111.
- Griffiths-Jones, S., R. J. Grocock, S. V. Dongen, A. Bateman, et A. J. Enright. 2006. "miRBase: microRNA sequences, targets and gene nomenclature". *Nucleic Acids Research*, Vol. 34, D140-D144.
- Gross, L. 1951. "Spontaneous leukemia developing in C3H mice following inoculation in infancy, with AK-leukemic extracts, or AK-embryos". *Proceedings of Social Experimental Biological Medecin*, Vol. 76, p. 27-32.
- Gu, Y., T. Nakamura, H. Alder, R. Prasad, O. Canaani, G. Cimino, C.M. Croce, et E. Canaani. 1992. "The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to *Drosophila* trithorax, to the AF-4 gene." *Cell*, Vol. 71, p. 701-708.

- Gu, H., and D.R. Schoenberg. 2003. "U2AF modulates poly(A) length control by the poly(A)-limiting element." *Nucleic Acids Research*, Vol. 31, p. 6264-6471.
- Haase, A.D., L. Jaskiewicz, H. Zhang, S. Laine, R. Sack, A. Gatignol, et C. W. Filipowicz. 2005. "TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing". *EMBO Reports*, Vol. 6, n°10, p. 961-967.
- Hadland, B.K., S.S. Huppert, J. Kanungo, Y. Xue, R. Jiang, T. Gridley, R.A. Conlon, A.M. Cheng, R. Kopan, et G.D. Longmore. 2004. "A requirement for Notch1 distinguishes 2 phases of definitive hematopoiesis during development." *Blood*, Vol. 104, p. 3097-3105.
- Hahn, W. C., et R. A. Weinberg. 2002. "Modelling the molecular circuitry of cancer". *Nature Reviews cancer*, Vol. 2, p. 331-341.
- Hanahan, D., et R.A. Weinberg. 2000. "The hallmarks of cancer." *Cell*, Vol. 100, p. 57-70.
- Hanna, Z., M. Janowski, P. Tremblay, X. Jiang, A. Milatovich, U. Francke, et P. Jolicoeur. 1993. "The Vin-1 gene, identified by provirus insertional mutagenesis, is the cyclin D2." *Oncogene*, Vol. 8, p. 1661-1666
- Harvey, J. J. 1964. "An unidentified virus which causes the rapid production of tumors in mice". *Nature*, Vol. 204, p. 1104.
- Hayward, W. S., B. G. Neel, et S. M. Astrin. 1981. "Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid Leukosis". *Nature*, Vol. 290, p. 475-480.
- He, Y., M. A. Brown, J. A. Rothnagel, N. A. Saunders, et R. Smith. 2005a. "Roles of heterogeneous nuclear ribonucleoproteins A and B in cell proliferation". *Journal of Cellular Science*, Vol. 118, p. 3173-3183.
- He, L., J. M. Thomson, M. T. Hemann, E. Hernando-Monge, D. Mu, S. Goodson, S. Powers, C. Cordon-Cardo, S. W. Lowe, G. J. Hannon, et S. M. Hammond. 2005b. "A microRNA polycistron as a potential human oncogene". *Nature*, Vol. 435, p. 828-833.
- Hoemann, C. D., N. Beaulieu, L. Girard, N. Rebai, et P. Jolicoeur. 2000. "Two Distinct Notch1 Mutant Alleles Are Involved in the Induction of T-Cell Leukemia in c-myc Transgenic Mice". *Molecular and Cellular Biology*, Vol. 20, n°11, p. 3831-3842.
- Hollander, M.C., I. Alamo, et A.J. Fornace Jr. 1996. "A novel DNA damage-inducible transcript, gadd7, inhibits cell growth, but lacks a protein product." *Nucleic Acids Research*, Vol. 24, p. 1589-1593.
- Hope, T.J. 1999. "The ins and outs of HIV Rev." *Arch. Biochem. Biophys*, Vol. 365, p. 186-191. Hornstein. E., et N. Shomron. 2006. "Canalization of development by microRNAs". *Nature Genetics*, Vol. 38, S20-S24.

- Hope, T.J. 1999. "The ins and outs of HIV Rev." *Arch. Biochem. Biophys.*, Vol. 365, p. 186-191.
- Hornstein, E., et N. Shomron. 2006. "Canalization of development by microRNAs". *Nature Genetics*, Vol. 38, S20-S24.
- Hornstein, E., et N. Shomron. 2006. "Canalization of development by microRNAs". *Nature Genetics*, Vol. 38, S20-S24.
- Hossain, A., M.T. Kuo, et G.F. Saunders. 2006. "Mir-17-5p Regulates Breast Cancer Cell Proliferation by Inhibiting Translation of *AIB1* mRNA." *Molecular and Cellular Biology*, Vol. 26, p. 8191-8201.
- Hutchison, S., C. LeBel, M. Blanchette, et B. Chabot. 2002. "Distinct Sets of Adjacent Heterogeneous Nuclear Ribonucleoprotein (hnRNP) A1/A2 Binding Sites Control 5' Splice Site Selection in the hnRNP A1 mRNA Precursor". *The Journal of Biological Chemistry*, Vol. 277, n°33, p. 29745-29752.
- Hwang, H. C., C. P. Martins, Y. Bronkhorst, E. Randel, A. Berns, M. Fero, et B. E. Clurman. 2002. "Identification of oncogenes collaborating with p27<sup>Kip1</sup> loss by insertional mutagenesis and high-throughput insertion site analysis". *Proceedings of National Academy of Science*, Vol. 99, n°17, p. 11293-11298.
- Ikegawa, S., M. Isomura, Y. Koshizuka, et Y. Nakamura. 1999. "Cloning and characterisation of ASH2L and Ash21, human and mouse homologs of the drosophila ash2 gene." *Cytogenetic and Cellular Genetic*, Vol. 84, p. 167-172.
- Ishikawa, F., M. J. Matunis, G. Dreyfuss, et T. R. Cech. 1993. "Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)*n*". *Molecular and Cellular Biology*, Vol. 13, p. 4301-4310.
- Johansson, F. K., J. Brodd, C. Eklof, M. Ferletta, G. Hesselager, C. F. Tiger, L. Uhrbom, et B. Westermarck. 2004. "Identification of candidate cancer-causing genes in mouse brain tumors by retroviral tagging". *Proceedings of National Academy of Science USA*, Vol. 101, n°31, p. 11334-11337.
- Johansson, F. K., H. Goransson, et B. Westermarck. 2005. "Expression analysis of genes involved in brain tumor progression driven by retroviral insertional mutagenesis in mice". *Oncogene*, Vol. 24, n°24, p. 3896-3905.
- Johnson, J. M., J. Castle, P. Garrett-Engele, Z. Kan, P. M. Loerch, C. D. Armour, R. Santos, E. E. Schadt, R. Stoughton, et D. D. Shoemaker. 2003. "Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays". *Science*, Vol. 302, p. 2141-2144.
- Johnson, S.M., H. Grosshans, J. Shingara, M. Byrom, R. Jarvis, A. Cheng, E. Labourier, K.L. Reinert, D. Brown, et F.J. Slack. 2005. "*RAS* Is Regulated by the *let-7* MicroRNA Family." *Cell*, Vol. 120, p. 635-647.

- Jonkers, J., et A. Berns. 1996. "Retroviral insertional mutagenesis as a strategy to identify cancer genes". *Biochemica and Biophysica Acta*, Vol. 1287, p. 29-57.
- Jonstrup, S.P., J. Koch, et J. Kjems. 2006. "A microRNA detection system based on padlock probes and rolling circle amplification." *RNA*, Vol. 12, p. 1747-1752.
- Joosten, M., Y. Vankan-Berkhoudt, M. Tas, M. Lunghi, Y. Jenniskens, E. Parganas, P. J. Valk, B. Lowenberg, E. van den Akker, et R. Delwel. 2002. "Large-scale identification of novel potential disease loci in mouse leukemia applying an improved strategy for cloning common virus integration sites". *Oncogene*, Vol. 21, n°47, p. 7247-7255.
- Kaiser, J. 2003. "Gene therapy. Seeking the cause of induced leukemias in X-SCID trial". *Science*, Vol. 299, p. 495.
- Kaplan, H.S. 1967. "On the natural history of the murine leukemias: presidential address." *Cancer Research*, Vol. 27, p. 1325-1340.
- Kasyapa, C. S., P. Kunapuli, et J. K. Cowell. 2005. "Mass spectroscopy identifies the splicing-associated proteins, PSF, hnRNP H3, hnRNP A2/B1, and TLS/FUS as interacting partners of the ZNF198 protein associated with rearrangement in myeloproliferative disease". *Experimental Cell Research*, Vol. 309, n°1, p. 78-85.
- Khvorova, A., A. Reynolds et S.D. Jayasena. 2003. "Functional siRNAs and miRNAs exhibit strand bias." *Cell*, Vol. 115, p. 209–216.
- Kim, R., A. Trubetskoy, T. Suzuki, N. A. Jenkins, N. G. Copeland, and J. Lenz. 2003. "Genome based identification of cancer genes by proviral tagging in mouse retrovirus-induced T-cell lymphomas". *Journal of Virology*, Vol. 77, p. 2056–2062.
- Kim, V. N., et J. W. Namb. 2006. "Genomics of microRNA". *Trends in Genetics*, Vol. 22, issue 3, p. 165-173.
- Kluiver, J., S. Poppema, D. de Jong, T. Blokzijl, G. Harms, S. Jacobs, B. J. Kroesen, et A. van den Berg. 2005. "BIC and miR-155 are highly expressed in Hodgkin, primary mediastinal and diffuse large B cell lymphomas". *Journal of Pathology*, Vol. 207, n°2, p. 243-249.
- Kozu, T., B. Henrich, et K. P. Schafer. 1995. "Structure and expression of the gene (HNRPA2B1) encoding the human hnRNP protein A2/B1". *Genomics*, Vol. 25, p. 365–371.
- Kramer, A. 1996. "The structure and function of proteins involved in mammalian pre-mRNA splicing". *Annual Review of Biochemistry*, Vol. 65, p. 367–409.
- Krause, M. O. 1996. "Chromatin structure and function: the heretical path to an RNA transcription factor". *Biochemistry and Cellular Biology*, Vol. 74, p. 623–632.

- Krecic, A. M., et M. S. Swanson. 1999. "hnRNP complexes: composition, structure, and function." *Current Opinion in Cell Biology*, Vol. 11, n°3, p. 363-371.
- Krek, A., D. Grun, M.N. Poy, R. Wolf, L. Rosenberg, E.J. Epstein, P. MacMenamin, I. da Piedade, K.C. Gunsalus, M. Stoffel, et N. Rajewsky. 2005. "Combinatorial microRNA target predictions." *Nature Genetics*, Vol. 37, p. 495-500.
- Kuhn, J., et S. Binder. 2002. "RT-PCR analysis of 5' and 3'-end-ligated mRNAs identifies the extremities of *cox2* transcripts in pea mitochondria." *Nucleic Acids Research*, Vol. 30, p. 439-446.
- Lagos-Quintana, M., R. Rauhut, W. Lendeckel, et T. Tuschl. 2001. "Identification of novel genes coding for small expressed RNAs". *Science*, Vol. 294, n°5543, p. 853–858.
- Lagos-Quintana, M., R. Rauhut, J. Meyer, A. Borkhardt, et T. Tuschl. 2003. "New microRNAs from mouse and human". *RNA*, Vol. 9, p. 175–179.
- Lai, E. C., P. Tomancak, R. W. Williams, et G. M. Rubin. 2003. "Computational identification of Drosophila microRNA genes". *Genome Biology*, Vol. 4, R42, p. 1–20.
- Lambert, J., D. Bergeron, C.A. Kozak, et E. Rassart. 1999. "Identification of a common site of provirus integration in radiation leukemia virus-induced T-cell lymphomas in mice." *Virology*, Vol. 264, p. 181-186.
- Landais, S., R. Quantin, et E. Rassart. 2005. "RadLV common integration at the *Kis2* locus : Simultaneous overexpression of a novel non-coding RNA and of the proximal *Phf6* gene." *Journal of Virology*, Vol. 79, n°17, p. 11443–11456.
- Landais, S., S. Landry, P. Legault, et E. Rassart. 2007. "Oncogenic potential of the miR-106-363 cluster and its implication in human T-cell leukemia". *Cancer Research*, Vol. 15, 67(12) p. 5699-5707.
- Landsberg, M. J., K. Moran-Jones, et R. Smith. 2006. "Molecular Recognition of an RNA Trafficking Element by Heterogeneous Nuclear Ribonucleoprotein A2". *Biochemistry*, Vol. 45, p. 3943-3951.
- Lanz, R. B., N. J. McKenna, S. A. Onate, U. Albrecht, J. Wong, S. Y. Tsai, M. J. Tsai, et B. W. O'Malley. 1999. "A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex". *Cell*, Vol. 97, p. 17–27.
- Lau, N. C., L. P. Lim, E. G. Weinstein, et D. P. Bartel. 2001. "An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*". *Science*, Vol. 294, n°5543, p. 858–862.
- Lazo, P. A., J. S. Lee, et P. N. Tshichlis. 1990. "Long-distance activation of the Myc proto-oncogene by provirus insertion in Mlvi-1 or Mlvi-4 in rat T-cell lymphomas". *Proceedings of National Academy of Science USA*, Vol. 87, p. 170-173.

- Lee, R. C., et V. Ambros. 2001. "An extensive class of small RNAs in *Caenorhabditis elegans*". *Science*. Vol. 294, n°543, p. 862–864.
- Lee, R. C., R. L. Feinbaum, et V. Ambros. 1993. "The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*". *Cell*, Vol. 75, n°5, p. 843–854.
- Lewis, B. P., I. H. Shih, M. W. Jones-Rhoades, D. P. Bartel, et C. B. Burge. 2003. "Prediction of mammalian microRNA targets". *Cell*, Vol. 115, p. 787–798.
- Lewis, B.P., C.B. Burge, et D.P. Bartel. 2005. "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets." *Cell*, Vol. 120, p. 15–20.
- Li, J., H. Shen, K. L. Himmel, A. J. Dupuy, D. A. Largaespada, T. Nakamura, J. D. Jr Shaughnessy, N. A. Jenkins, et N. G. Copeland. 1999. "Leukaemia disease genes: large-scale cloning and pathway predictions". *Nature Genetics*, Vol. 23, p. 348–353.
- Lieberman, M., O. Niwa, A. Declève, et H.S. Kaplan. 1973. "Continuous propagation of radiation leukemia virus on a C57BL mouse embryo fibroblast line with attenuation of leukemogenic activity." *Proceedings of National Academy of Science USA*, Vol. 70, p. 1250–1253.
- Lieberman, M., A. Declève, E.P. Gellmann, et H.S. Kaplan. 1977. "Biological and serological characterization of the C-type RNA viruses isolated from C57BL/Ka strain of mice.II. Induction and propagation of the isolates in radiation-induced leukemogenesis and related viruses," p.231–246. In J.F. Duplan (ed.), *INSERM Symposium no.4*, Paris. France.
- Lieberman, M., A. Declève, P. Ricciardi-Castagnoli, J. Bonier, O. J. Finn, et H. S. Kaplan. 1979. "Establishment, characterization and virus expression of cell lines derived from radiation-and virus-induced lymphomas of C57BL/Ka mice". *International Journal of Cancer*, Vol. 24, p. 168–177.
- Lim, L. P., N. C. Lau, E. G. Weinstein, A. Abdelhakim, S. Yekta, M. W. Rhoades, C. B. Burge, et D. P. Bartel. 2003. "The microRNAs of *Caenorhabditis elegans*". *Genes and Development*, Vol. 17, p 991–1008 a.
- Lim, L.P., M. E. Glasner, S. Yekta, C. B. Burge, et D. P. Bartel. 2003. "Vertebrate microRNA genes". *Science*, Vol. 299, p. 1540 b.
- Lim, L. P., N. C. Lau, P. Garrett-Engele, A. Grimson, J. M. Schelter, J. Castle, D. P. Bartel, P. S. Linsley, et J. M. Johnson. 2005. "Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs ". *Nature*, Vol. 433, p. 769–773.
- Liu, C. G., G. A. Calin, B. Meloon, N. Gamliel, C. Sevignani, M. Ferracin, C. D. Dumitru, M. Shimizu, S. Zupo, M. Dono, H. Alder, F. Bullrich, M. Negrini, et C. M. Croce. 2004. "An oligonucleotide microchip for genome-wide microRNA profiling

- Lower, K.M., G. Turner, B.A. Kerr, K.D. Mathews, M.A. Shaw, A.K. Gedeon, S. Schelley, H.E. Hoyme, S.M. White, M.B. Delatycki, A.K. Lampe, J. Clayton-Smith, H. Stewart, C.M. van Ravenswaay, B.B. de Vries, B. Cox, M. Grompe, S. Ross, P. Thomas, J.C. Mulley, et J.Gecz. 2002. "Mutations in PHF6 are associated with Börjeson-Forssman-Lehmann syndrome." *Nature Genetics*, Vol. 32, p. 661-665.
- Lu, J., G. Getz, E. A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B. L. Ebert, R. H. Mak, A. A. Ferrando, J. R. Downing, T. Jacks, H. R. Horvitz, et T. R. Golub. 2005. "MicroRNA expression profiles classify human cancers". *Nature*, Vol. 435, p. 834-838.
- Lund, A.H., G. Turner, A. Trubetskoy, E. Verhoeven, E. Wientjens, D. Hulsman, R. Russel, R.A. DePinho, J. Lenz, et M. Van Lohuizen. 2002. "Genome wide retroviral insertional tagging of genes involved in cancer in Cdkn2a-deficient mice." *Nature Genetics*, Vol. 32, p. 160-165.
- Luo, M. J., et R. Reed. 1999. "Splicing is required for rapid and efficient mRNA export in metazoans". *Proceedings of National Academy of Science USA*, Vol. 96, n°26, p. 14937-14942.
- Lyngso, C., G. Bouteiller, C.K. Damgaard, D. Ryom, S. Sanchez-Munoz, P.L. Norby, B.J. Bonven, et P. Jorgensen. 2000. "Interaction between the transcription factor SPBP and the positive cofactor RNF4 : an interplay between protein binding zing fingers." *Journal of Biological Chemistry*, Vol. 275, p. 26144-26149.
- Ma, A. S. W., K. Moran-Jones, J. Shan, T. P. Munro, M. J. Snee, K. S. Hoek , et R. Smith. 2002. " Heterogeneous Nuclear Ribonucleoprotein A3, a Novel RNA Trafficking Response Element-binding Protein". *The Journal of Biological Chemistry*, Vol. 277, n°20, p. 18010-18020.
- Maillard, I., A.P. Weng, A.C. Carpenter, C.G. Rodriguez, H. Sai, L. Xu, D. Allman, J.C. Aster, et W.S. Pear. 2004. "Mastermind critically regulates Notch-mediated lymphoid cell fate decisions." *Blood*, Vol. 104, p. 1696-1702.
- Markovtsov, V., J. M. Nikolic, J. A. Goldman, C. W. Turck, M-Y Chou, et D. L. Black. 2000. "Cooperative Assembly of an hnRNP Complex Induced by aTissue-Specific Homolog of Polypyrimidine Tract Binding Protein". *Molecular and Cellular Biology*, Vol. 20, n°20, p. 7463-7479.
- Maroney, P. A., Y. Yu, J. Fisher, et T. W. Nilsen. 2006. "Evidence that microRNAs are associated with translating messenger RNAs in human cells". *Nature structural and molecular biology*, Vol. 13, n°12, p. 1102-1107.
- Martin-Hernandez, J., A.B. Sorensen, et F.S. Pedersen. 2001. "Murine leukemia virus proviral insertions between the N-ras and unr genes in B-cell lymphoma DNA affect the expression of N-ras only." *Journal of Virology*, Vol. 75, :p. 11907-11912.



- Matlin, A. J., F. Clark, et C. W. J. Smith. 2005. "Understanding alternative splicing: towards a cellular code". *Nature Reviews Molecular Cell Biology*, Vol. 6, p. 386-398.
- Mattick, J. S. 2001. "Non-coding RNAs : the architects of eukaryotic complexity". *EMBO reports*, Vol. 21, n°11, p. 986-991.
- Mattick, J. S. 2003. "Challenging the dogma : the hidden layer of non-protein-coding RNAs in complex organism". *Bioessays*, Vol. 25. p. 930-939.
- Mayeda, A., S. H. Munroe, J. F. Cáceres, et A. R. Krainer. 1994. "Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins". *EMBO journal*, Vol. 13, p. 5483-5495.
- McKay, S., et H. Cooke. 1992. "hnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat TTAGGGn". *Nucleic Acids Research*, Vol. 20, p. 6461-6464.
- Meller, V.H., K.H. Wu, G. Roman, M.I. Kuroda, et R.L. Davis. 1997. "roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system." *Cell*, Vol. 88, p. 445-457.
- Merlino, G. T., Y. H. Xu, S. Ishii, A. J. Clark, K. Semba, K. Toyoshima, T. Yamamoto, et I. Pastan. 1984. "Amplification and enhanced expression of the epidermal growth factor receptor gene in A431 human carcinoma cells". *Science* Vol. 224, p. 417-419.
- Mikkers, H., J. Allen, P. Knipscheer, L. Romeyn, A. Hart, E. Vink et A. Berns. 2002. "High-throughput retroviral tagging to identify components of specific signaling pathways in cancer". *Nature Genetics*, Vol. 32, p. 153-159.
- Min, H., C. W. Turck, J. M. Nikolic, et D. L. Black. 1997. "A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer". *Genes and Development*, Vol. 11, n°8, p. 1023-1036.
- Modafferi, E. F., et D. L. Black. 1997. "A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon". *Molecular and Cellular Biology*, Vol. 17, p. 6537-6545.
- Moloney, J. B. 1960. "Biological studies on a lymphoid-leukemia virus extracted from sarcoma 37. I. Origin and introductory investigations". *Journal of National Cancer Institut*, Vol. 24, p. 933-951.
- Moran-Jones, K., L. McLure, D. Kennedy, R. R. Reddel, S. Sara, et R. Smith. 2005. "hnRNP A2, a potential ssDNA/RNA molecular adapter at the telomere". *Nucleic Acids Research*, Vol. 33, p. 486-496.
- Mount, S.M., C. Burks, G. Hert, G. D. Stormo, O. White, et C. Fields. 1992. "Splicing signals in Drosophila: intron size, information content, and consensus sequences". *Nucleic Acids Research*, Vol. 20, n°16, p. 4255-4262.

- Mourelatos, Z., J. Dostie, S. Paushkin, A. Sharma, B. Charroux, L. Abel, J. Rappsilber, M. Manu, et G. Dreyfuss. 2002. "miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs". *Genes and Development*, Vol. 16, p. 720-728.
- Munro, T. P., R. J. Magee, G. J. Kidd, J. H. Carson, E. Barbarese, L. M. Smith, et R. Smith. 1999. "Mutational analysis of a heterogeneous nuclear ribonucleoprotein A2 response element for RNA trafficking". *The Journal of Biological Chemistry*, Vol. 274, p. 34389-34395.
- Nelson, P. T., D. A. Baldwin, L. M. Scearce, J. C. Oberholtzer, J. W. Tobias, et Z. Mourelatos. 2004. "Microarray-based, high-throughput gene expression profiling of microRNAs". *Nature Methods*, Vol. 1, p. 155-161.
- Obernosterer, G., P. J. F. Leuschner, M. Alenius, et J. Martinez. 2006. "Post-transcriptional regulation of microRNA expression". *RNA*, Vol. 12, p. 1161-1167.
- O'Donnell, K. A., E. A. Wentzel, K. I. Zeller, C. V. Dang, et J. T. Mendell. 2005. "c-Myc-regulated microRNAs modulate E2F1 expression". *Nature*, Vol. 435, p. 839-843 (2005).
- Olsen, P. H., et V. Ambros. 1999. "The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation". *Developmental Biology*, Vol. 216, p. 671-680.
- Parker, J. S., S. M. Roe, et D. Barford. 2005. "Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex". *Nature*, Vol. 434, p. 663-666.
- Patry, C., L. Bouchard, P. Labrecque, D. Gendron, B. Lemieux, J. Toutant, E. Lapointe, R. Wellinger, et B. Chabot. "Small interfering RNA-mediated reduction in heterogeneous nuclear ribonucleoprotein A1/A2 proteins induces apoptosis in human cancer cells but not in normal mortal cell lines". *Cancer Research*, Vol. 63, n°22, p. 7679-7688.
- Patton, J. G., E. B. Porro, J. Galceran, P. Tempst, et B. Nadal-Ginard. 1993. "Cloning and characterization of PSF, a novel pre-mRNA splicing factor". *Genes and Development*, Vol. 7, p. 393-406.
- Pekarsky, Y., U. Santanam, A. Cimmino, A. Palamarchuk, A. Efanov, V. Maximov, S. Volinia, H. Alder, C. G. Liu, L. Rassenti, G. A. Calin, J. P. Hagan, T. Kipps, et C. M. Croce. 2006. "Tcl1 expression in chronic lymphocytic leukemia is regulated by miR-29 and miR-181". *Cancer Research*, Vol. 66, n°24, p. 11590-11593.
- Perez, I., C. H. Lin, J. G. McAfee, et J. G. Patton. 1997. "Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo". *RNA*, Vol. 3, n°7, p. 764-778.
- Peters, J. 2000. "Imprinting: silently crossing the boundary". *Genome Biology*, Vol. 1, n°5, reviews 1028.

- Qiao, Y. L., M. S. Tockman, L. Li, Y. S. Erozan, S. X. Yao, M. J. Barrett, W. H. Zhou, C. A. Giffen, X. C. Luo, et P. R. Taylor. 1997. "A case-cohort study of an early biomarker of lung cancer in a screening cohort of Yunnan tin miners in China". *Cancer Epidemiology*, Vol. 6, p. 893–900.
- Rajewsky, N. 2006. "microRNA target predictions in animals". *Nature Genetics*, Vol. 38, S8-13.
- Rappsilber, J., U. Rider, A. I. Lamond, et M. Mann. 2002. "Large-scale proteomic analysis of the human spliceosome". *Genome Research*, Vol. 12, n°8, p. 31-45.
- Rassart, E., P. Sankar-Mistry, G. Lemay, L. Desgroseillers, et P. Jolicoeur. 1983. "New class of leukemogenic ecotropic recombinant murine leukemia virus isolated from radiation induced thymomas of C57BL/6 mice". *Journal of virology*, Vol. 45, p. 565-575.
- Rassart, E., M. Shang, Y. Boie, et P. Jolicoeur. 1986. "Studies on emerging Radiation leukemia virus variants in C57BL/Ka mice". *Journal of Virology*, Vol. 58, p. 96-106.
- Reinhart, B. J., F. J. Slack, M. Basson, A. E. Pasquinelli, J. C. Bettinger, A. E. Rougvie, H. R. Horvitz, G. Ruvkun. 2000. "The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis Elegans*". *Nature*, Vol. 403, p. 901-906.
- Rooke, N., V. Markovtsov, E. Cagavi, et D. L. Black. 2003. "Roles for SR proteins and hnRNP A1 in the regulation of c-src exon N1". *Molecular and Cellular Biology*, Vol. 23, n°6, p. 1874-1884.
- Rosonina, E., J. Y. Y. Ip, J. A. Calarco, M. A. Bakowski, A. Emili, S. McCracken, P. Tucker, C. J. Ingles, et B. J. Blencowe. 2005. "Role for PSF in Mediating Transcriptional Activator-Dependent Stimulation of Pre-mRNA Processing In Vivo". *Molecular and Cellular Biology*, Vol. 25, n°15, p. 6734–6746.
- Rous, P. 1910. *Journal of Experimental Medicine*, vol. 12, p. 696–705.
- Rous, P. 1911. *Journal of Experimental Medicine*, vol. 13, p. 397–411.
- Sambrook, J., et D.W Russell. 2001. "Molecular cloning : a laboratory manual", 3<sup>rd</sup> ed. *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, N.Y.
- Sanchez-Elsner, T., D. Gou, E. Kremmer, et F. Sauer. 2006. "Noncoding RNAs of Trithorax Response Elements Recruit *Drosophila* Ash1 to Ultrabithorax". *Science*, Vol. 311, n°5764, p. 1118 – 1123.
- Schwab, R., J. F. Palatnik, M. Riester, C. Schommer, M. Schmid, et D. Weigel. 2005. "Specific effects of microRNAs on the plant transcriptome". *Developmental Cell*, Vol. 8, p. 517–527.

- Schwarz, D.S., G. Hutvagner, T. Du, Z. Xu, N. Aronin et P.D. Zamore. 2003. "Asymmetry in the assembly of the RNAi enzyme complex." *Cell*, Vol. 115, p. 199–208. *Science*, Vol. 265, p. 1706-1709.
- Selby, C. C., C. E. Grey, S. Lichtenberg, C. Friend, A. E. Moore, et J. J Biesele. 1954. "Submicroscopic cytoplasmic particles occasionally found in the Ehrlich mouse ascites tumor"..*Cancer Research*, Vol. 14, p. 790-794.
- Selten, G., H. T. Cuypers, et A. Berns. 1985. "Proviral activation of the putative oncogene *Pim-1* in MuLV induced T-cell lymphomas". *The EMBO Journal*, Vol. 4, n°7, p. 1793-1798.
- Selten, G., H.T. Cuypers, M. Zijlstra, C. Melief et A. Berns. 1984. "Involvement of c-Myc in MuLV-induced T cell lymphomas in mice : frequency and mechanisms of activation." *EMBO*, Vol. 3, p. 3215-3222.
- Sempere, L. F., N. S. Sokol, E. B. Dubrovsky, E. M. Berger, et V. Ambros. 2003. "Temporal regulation of microRNA expression in *Drosophila melanogaster* mediated by hormonal signals and broad-Complex gene activity". *Developmental Biology*, Vol. 259, p. 9–18.
- Seggerson, K., L. Tang, et E. G. Moss. 2002. "Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene *lin-28* after translation initiation". *Developmental Biology*, Vol. 243, p. 215–225.
- Shen, H., et M. R. Green. 2004. "A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly". *Molecular Cell*, Vol. 16, p. 363–373.
- Shibata, S. et J.T. Lee. 2003. "Characterisation and quantitation of differential *Tsix* transcripts : implication for *Tsix* function." *Human Molecular Genetic*, Vol. 12, p. 125-136.
- Shin, M. S., T. N. Fredrickson, J. W. Hartley, T. Suzuki, K. Agaki, et H. C. Morse III. 2004. "High-Throughput Retroviral Tagging for Identification of Genes Involved in Initiation and Progression of Mouse Splenic Marginal Zone Lymphomas". *Cancer Research*, Vol. 64, p. 4419-4427.
- Singh, R., J. Valcarcel, et M. R. Green. 1995. "Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins". *Science*, Vol. 268, p. 1173-1176.
- Singh, R., J. Valcarcel, et M. R. Green. 1995. "Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins". *Science*, Vol. 268, p. 1173-1176.

- Singh, R., et J. Valcarcel. 2005. "Building specificity with non specific RNA binding proteins". *Nature Structural and Molecular Biology*, Vol. 12, p. 645-654.
- Sørensen, A. B., M. Duch, H. W. Amtoft, P. Jørgensen et F. S. Pedersen. 1996. "Sequence tags of provirus integration sites in dans of tumors induced by the murine retrovirus SL3-3". *Journal of Virology*, Vol. 70, n°6, p. 4063-4070.
- Stark, A., J. Brennecke, N. Bushati, R. B. Russell, et S. M. Cohen. 2005. "Animal MicroRNAs Confer Robustness to Gene Expression and Have a Significant Impact on 3'UTR Evolution". *Cell*, Vol. 123, p. 1133-1146.
- Stéhelin, D., H. E. Varmus, J. M. Bishop et P. K. Vugt. 1976. "DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA". *Nature*, Vol. 260, p. 170-173.
- Suzuki, T., H. Shen, K. Akagi, H. C. Morse III, J. D. Malley, D. Q. Naiman, N. A. Jenkins et N. G. Copeland. 2002. "New genes involved in cancer identified by retroviral tagging". *Nature Genetics*, Vol. 32, p. 166-174.
- Sylvestre, Y., V. De Guire, E. Querido, U. K. Mukhopadhyay, V. Bourdeau, F. Major, G. Ferbeyre, et P. Chartrand. 2006. "An E2F/miR-20a auto-regulatory feed-back loop". *Journal of biological chemistry*, advanced online publication.
- Takamizawa, J., H. Konishi, K. Yanagisawa, S. Tomida, H. Osada, H. Endoh, T. Harano, Y. Yatabe, M. Nagino, Y. Nimura, T. Mitsudomi, et T. Takahashi. 2004. "Reduced Expression of the *let-7* MicroRNAs in Human Lung Cancers in Association with Shortened Postoperative Survival." *Cancer Research*, Vol. 64, p. 3753-3756.
- Tam, W., D. Ben-Yehuda, et W. S. Hayward. 1997. "bic, a novel gene activated by proviral insertions in avian leukosis virus-induced lymphomas, is likely to function through its noncoding RNA". *Molecular and Cellular Biology*, Vol. 17, n°3, p. 1490-1502.
- Tam, W., S. H. Hughes, W. S. Hayward, et P. Besmer. 2002. "Avian bic, a gene isolated from a common retroviral site in avian leukosis virus-induced lymphomas that encodes a noncoding RNA, cooperates with c-myc in lymphomagenesis and erythroleukemogenesis". *Journal of Virology*, Vol. 76, p. 4275-4286.
- Thomson, J. M., M. Newman, J. S. Parker, E. M. Morin-Kensicki, T. Wright, et S. M. Hammond. 2006. "Extensive post-transcriptional regulation of microRNAs and its implications for cancer". *Genes and Development*, Vol. 20, p. 2202-2207.
- Tkachuk, D.C., S. Kohler, et M.L. Cleary. 1992. "Involvement of a homologue of *Drosophila* trithorax by 11q23 chromosomal translocations in acute leukemias." *Cell*, Vol. 71, p. 691-700.
- Tockman, M. S., J. L. Mulshine, S. Piantadosi, Y. S. Erozan, P. K. Gupta, J. C. Ruckdeschel,

- P. R. Taylor, T. Zhukov, W. H. Zhou, Y. L. Qiao, et S. X. Yao. 1997. "Prospective detection of preclinical lung cancer: results from two studies of heterogeneous nuclear ribonucleoprotein A2/B1 overexpression". *Clinical Cancer Research*, Vol. 3, p. 2237–2246.
- Tremblay, P. J., C. A. Kosak, et P. Jolicoeur. 1992. "Identification of a novel gene, Vin-1, in murine Leukemia Virus-induced T-cell leukemia by provirus insertional mutagenesis". *Journal of virology*, Vol. 66, p. 1344-1353.
- Ushigome, M., T. Ubagai, H. Fukuda, N. Tsuchiya, T. Sugimura, J. Takatsuka, et H. Nakagama. 2005. "Up-regulation of hnRNP A1 gene in sporadic human colorectal cancers". *International Journal of Oncology*, Vol. 26, n°3, p. 635-640.
- van der Lugt, N. M. T., J. Domen, E. Verhoeven, K. Linders, H. van der Gulden, J. Allen, et A. Berns. 1995. "Proviral tagging in Eμ-myc transgenic mice lacking the Pim-1 proto-oncogene leads to compensatory activation of Pim-2". *The EMBO Journal*, Vol. 14, n°11, p. 2536-2544.
- van Lohuizen, M., S. Verbeek, P. Krimpenfort, J. Domen, C. Saris, T. Radaszkiewicz, et A. Berns. 1989. "Predisposition to lymphomagenesis in pim-1 transgenic mice: cooperation with c-myc and N-myc in murine leukemia virus-induced tumors. *Cell*, Vol. 56, p. 673-682.
- van Lohuizen, M., S. Verbeek, B. Scheijen, E. Wientjens, H. van der Gulden, et A. Berns. 1991. "Identification of cooperating oncogenes in Eμ-myc transgenic mice by provirus tagging". *Cell*, Vol. 65, p. 737-752.
- Vella, M. C., E. Y. Choi, S. Y. Lin, K. Reinert K, et F. J. Slack. 2004. "The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR". *Genes and Development*, Vol. 15, n°2, p. 132-137.
- Volinia, S., G. A. Calin, C. G. Liu, S. Ambs, A. Cimmino, F. Petrocca, R. Visone R, M. Iorio, C. Roldo, M. Ferracin, R. L. Prueitt, N. Yanaihara, G. Lanza, A. Scarpa, A. Vecchione, M. Negrini, C. C. Harris, et C. M. Croce. 2006. "A microRNA expression signature of human solid tumors define cancer gene targets". *Proceedings of National Academy of Science USA*, Vol. 103, p. 2257–2261.
- Voorhoeve, P.M., C. Le Sage, M. Schrier, A.J. Gillis, H. Stoop, R. Nagel, Y.P. Liu, J. van Duijse, J. Drost, A. Griekspoor, E. Zlotorynski, N. Yabuta, G. De Vita, H. Nojima, L.H. Looijenga, et R. Agami. 2006. "A Genetic Screen Implicates miRNA-372 and miRNA-373 As Oncogenes in Testicular Germ Cell Tumors." *Cell*, Vol. 124, p. 1169–1181.
- Wallace, A.M., D. Brinda, S.E. Ravnik, V. Tonk, N.A. Jenkins, D.J. Gilbert, N.G. Copeland, et C.C. MacDonald. 1999. "Two distinct forms of the 64,000 Mr protein of the cleavage stimulation factor are expressed in mouse male germ cells." *Proceedings of National Academy of Science USA*, Vol. 96, p. 6763-6768.

- Wang, Y., D.R. Crawford, et K.J.A. Davies. 1996. "*adapt33*, a novel oxidant-inducible RNA from hamster HA-1 cells." *Arch. Bioch. Bioph*, Vol. 332, p. 255-260.
- Wang, C.L., B.B. Wang, G. Bartha, L. Li, N. Channa, M. Klinger, N. Killeen, et M. Wabl. 2006. "Activation of an oncogenic microRNA cistron by provirus integration." *Proceedings of National Academy of Science USA*, Vol. 103, p. 18680–18684.
- Wang, Z., X. Xiao, E. V. Nostrand, et C. B. Burge. 2006. "General and Specific Functions of Exonic Splicing Silencers in Splicing Control". *Molecular cell*, Vol. 23, p. 61-70.
- Weng, A.P., A.A. Ferrando, W. Lee, J.P. 4th Morris, L.B. Silverman, C. Sanchez-Irizarry, S.C. Blacklow, A.T. Look, et J.C. Aster. 2004. "Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia." *Science*, Vol. 306, p. 269-271.
- Wightman, B., T. R. Burglin , J. Gatto , P. Arasu , et G. Ruvkun . 1991. "Negative regulatory sequences in the *lin-14* 3'-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development". *Genes and Development*, Vol. 5, n°10, p. 1813–1824.
- Wightman, B., I. Ha , et G. Ruvkun. 1993. "Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*". *Cell*. Vol. 75, n°5, p. 855–862.
- Williams, K. R., K. L. Stone, M. B. LoPresti, B. M. Merrill, et S. R. Planck. 1985. "Amino acid sequence of the UPI calf thymus helix-destabilizing protein and its homology to an analogous protein from mouse myeloma". *Proceedings of National Academy of Science USA*, Vol. 82, n°17, p. 5666-5670.
- Woods, K., J. M. Thomson, et S. M. Hammond. 2006. "Direct regulation of an oncogenic microRNA cluster by E2F transcription factors". *Journal of biological chemistry*, advanced online publication.
- Wu, X., Y. Li, B. Crise, et S. M. Burgess. 2003. "Transcription start regions in the human genome are favored targets for MLV integration". *Science*, Vol. 300, p. 1749-51.
- Xu, N., C.Y. Chen, et A.B. Shyu. 1997. "Modulation of the fate of cytoplasmic mRNA by AU-rich elements: key sequence features controlling mRNA deadenylation and decay." *Molecular and Cellular Biology*, Vol. 17, p. 4611-4621.
- Yanagawa S, Lee JS, Kakimi K, Matsuda Y, Honjo T, et A. Ishimoto. 2000. "Identification of Notch1 as a frequent target for provirus insertional mutagenesis in T-cell lymphomas induced by leukemogenic mutants of mouse mammary tumor virus." *Journal of Virology*, Vol. 74, p. 9786-9791.

- Yekta, S., I. H. Shih, et D. P. Bartel,. 2004. "MicroRNA-directed cleavage of HOXB8 mRNA". *Science*, Vol. 304, p. 594–596.
- Yu, J., F. Wang, G. H. Yang, F. L. Wang, Y. N Ma, Z. W. Du, et J. W. Zhang. 2006. Human microRNA clusters: Genomic organization and expression profile in leukemia cell lines. *Biochemistry and Biophysical Research Commun*, Vol. 349, n°1, p. 59-68.
- Zhang, M. Q., et T. G. Marr. 1994. "Fission yeast gene structure and recognition". *Nucleic Acids Research*, Vol. 22, n°9, p. 1750-1759.
- Zhang, B., Q. Wang, et X. Pan. 2006a. "MicroRNAs and their regulatory roles in animals and plants". *Journal of cellular physiology*, Vol. 210, issue 2, p. 279-289.
- Zhang, B., X. Pan, G. P. Cobb, et T. A. Anderson. 2006b. microRNAs as oncogenes and tumor suppressors. *Development Biology*, advanced publication on line.
- Zhou, Z., L. J. Licklider, S. P. Gygi, et R. Reed. 2002. "Comprehensive proteomic analysis of the human spliceosome". *Nature*, Vol. 419, p. 182-185.